(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) **EP 2 893 526 B1**

(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication and mention of the grant of the patent:
12.04.2017 Bulletin 2017/15

(21) Application number: **13770552.1**

(22) Date of filing: **10.09.2013**

(51) Int Cl.:
G10L 15/22 *(2006.01)*

(86) International application number:
PCT/US2013/058916

(87) International publication number:
WO 2014/040022 (13.03.2014 Gazette 2014/11)

(54) CONTEXT-SENSITIVE HANDLING OF INTERRUPTIONS BY INTELLIGENT DIGITAL ASSISTANTS

KONTEXTSENSITIVE HANDHABUNG VON UNTERBRECHUNGEN DURCH INTELLIGENTE DIGITALE ASSISTENTEN

GESTION SENSIBLE AU CONTEXTE D'INTERRUPTIONS PAR DES ASSISTANTS NUMÉRIQUES INTELLIGENTS

(84) Designated Contracting States:
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR

(30) Priority: **10.09.2012 US 201261699259 P**
09.09.2013 US 201314021972

(43) Date of publication of application:
15.07.2015 Bulletin 2015/29

(73) Proprietor: **Apple Inc.**
Cupertino, CA 95014 (US)

(72) Inventor: **VAN OS, Marcel**
Cupertino, CA 95014 (US)

(74) Representative: **Gillard, Matthew Paul et al**
Withers & Rogers LLP
4 More London Riverside
London SE1 2AU (GB)

(56) References cited:
US-A1- 2008 077 310
US-A1- 2009 171 578
US-B1- 6 208 932

Note: Within nine months of the publication of the mention of the grant of the European patent in the European Patent Bulletin, any person may give notice to the European Patent Office of opposition to that patent, in accordance with the Implementing Regulations. Notice of opposition shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European Patent Convention).

EP 2 893 526 B1

Description

TECHNICAL FIELD

[0001] The disclosed embodiments relate generally to digital assistants, and more specifically, to digital assistants that intelligently handle user-initiated and/or system-initiated interruptions based on the current context.

BACKGROUND

[0002] Just like human personal assistants, digital assistants or virtual assistants can perform requested tasks and provide requested advice, information, or services. An assistant's ability to fulfill a user's request is dependent on the assistant's correct comprehension of the request or instructions. Recent advances in natural language processing have enabled users to interact with digital assistants using natural language, in spoken or textual forms, rather than employing a conventional user interface (*e.g.,* menus or programmed commands). Such digital assistants can interpret the user's input to deduce the user's intent; translate the deduced intent into actionable tasks and parameters; execute operations or deploy services to perform the tasks; and produce outputs that are intelligible to the user. Ideally, the outputs produced by a digital assistant should fulfill the user's intent expressed during the natural language interaction between the user and the digital assistant.

[0003] The ability of a digital assistant system to produce satisfactory responses to user requests depends on the natural language processing, knowledge base, and artificial intelligence implemented by the system. A well-designed response procedure can improve a user's experience in interacting with the system and promote the user's confidence in the system's services and capabilities.

[0004] US2008/0077310A1 discloses an in-vehicle telematics system that arbitrates between providing spoken navigation prompts and carrying out another user interaction, such as allowing a user to place a telephone call.

SUMMARY

[0005] The invention is defined by the appended claims, to which reference should now be made. The following paragraphs of this Summary serve to highlight various features of some embodiments of the invention.

[0006] The embodiments disclosed herein provide methods, systems, computer readable storage medium and user interfaces for a digital assistant to intelligently and dynamically determine how to handle a user-initiated and/or system-initiated interruption to an existing task currently underway based on the current context. In some embodiments, the digital assistant evaluates the relative urgency between delivering an output associated with the existing task and delivering an output associated with the interruption, and determines how to prioritize the deliveries of the outputs based on the determined relative urgency. In some embodiments, the relative urgency is evaluated based on a number of relevant factors (*e.g.,* also referred to as "priority parameters") forming the current context. In some embodiments, the relative urgency is only evaluated and used to prioritize deliveries of respective outputs associated the existing task and the interruption when the digital assistant detects a conflict between the default delivery times and output channels for the two outputs. In some embodiments, based on the dynamically determined relative urgency, the digital assistant prioritizes the deliveries of the two outputs and carries out the deliveries according to the priorities in real-time. In some embodiments, in addition to the relative urgency, the digital assistant also evaluates how important it is for the outputs to be delivered at their respective default delivery times, i.e., the relative flexibility in delivery time for the outputs. In some embodiments, the digital assistant adjusts the delivery times for one or both of the outputs based on both the determined relative urgency and the determined relative flexibility between the two outputs.

[0007] Accordingly, some examples provide a method for operating a digital assistant, the method including, at a device including one or more processors and memory storing one or more programs: receiving a first speech input from a user; initiating a first information provision process in response to receipt of the first speech input, the first information provision process comprising preparing at least a first response and a second response to the first speech input; providing the first response to the user; after or concurrent with the provision of the first response to the user, but before provision of the second response to the user, detecting an event operable to initiate a second information provision process; initiating the second information provision process in response to detecting the event, the second information provision process comprising preparing at least a third response to the event; determining a relative urgency between the second response and the third response; and providing one of the second response and the third response to the user in an order based on the determined relative urgency.

[0008] In some embodiments, the first and second responses are two consecutive sub-responses of a series of discrete sub-responses to the first speech input. In some embodiments, the first information provision process further comprises providing the series of discrete sub-responses to the user one at a time over an extended period of time without requiring a further prompt from the user.

[0009] In some embodiments, provision of all of the series of discrete sub-responses terminates the first information provision process.

[0010] In some embodiments, the first speech input is a navigation request and the first response and the second response are two navigation instructions associated

with two different waypoints along a route prepared in response to the navigation request.

[0011] In some embodiments, the first speech input is a search request, and the first response and the second response are speech outputs reading two different search results retrieved in response to the search request.

[0012] In some embodiments, the first speech input is a list-reading request, and the first response is a speech output summarizing a list of information items or a subset thereof, or a speech output reading content of at least one of the list of information items. In some embodiments, the list-reading request is a request to read one of: a cooking recipe, a list of email messages, a list of search results, a list of instructions, a list of diagnostic procedures, a list of exercise routines, a list of calendar entries, a list of reminders, a list of navigation instructions, a list of voice mail messages, and a list of SMS messages.

[0013] In some embodiments, the first speech input is a request to establish a reminder to be triggered at a later time by occurrence of a specified triggering event, the first response is an acknowledgement to the first speech input, and the second response is an alert item to be delivered to the user at the later time.

[0014] In some embodiments, the event is receipt of a second speech input from the user. In some embodiments, the third response is a complete response to the second speech input. In some embodiments, the third response is an initial sub-response among a series of sub-responses to the second speech input to be provided to the user over an extended period of time.

[0015] In some embodiments, the event is occurrence of a trigger event for a previously established reminder. In some embodiments, the third response is a speech output providing content of the reminder.

[0016] In some embodiments, the event is arrival of a push-notification. In some embodiments, the third response is a speech output providing content of the push-notification.

[0017] In some embodiments, the event is receipt of a second speech input that does not alter validity of the second response.

[0018] In some embodiments, the event includes receipt of a speech input directed to the digital assistant and generation of an alert or reminder by an application or process currently controlled by the digital assistant.

[0019] In some embodiments, the first speech input is a first information request, the event is receipt of a second information request, and the second response and the third response are two speech outputs providing respective information requested by the first and second information requests.

[0020] In some embodiments, the first speech input is a request to establish a reminder to be delivered at a later time, the event is receipt of an information request, the second response is an alert item providing content of the established reminder, and the third response is a speech output providing information retrieved in response to the information request.

[0021] In some embodiments, the first speech input is an information request, the event is occurrence of a trigger event for a previously established reminder or arrival of a push-notification, the second response is a speech output providing information retrieved in response to the information request, and the third response is an alert item providing content of the previously established reminder or push notification.

[0022] According to the invention the method further includes: determining whether the digital assistant is currently operating in a hands-free mode or an eyes-free mode; and determining the relative urgency between the second response and the third response upon determining that the digital assistant is currently operating in the hands-free mode or eyes-free mode.

[0023] In some embodiments, the method further includes: detecting that the user is currently in motion; and invoking the hands-free mode or the eyes-free mode of the digital assistant upon detecting that the user is currently in motion.

[0024] In some embodiments, the method further includes: detecting that the user is currently performing one of the actions of: navigating a vehicle, walking, jogging, exercising, and operating a device or application not currently controlled by the digital assistant; and invoking the hands-free mode or the eyes-free mode upon said detection.

[0025] In some embodiments, the method further includes: determining whether the second response and the third response are suitable to be provided in parallel on different output channels of the device; and determining the relative urgency between the second and the third responses upon determining that the second response and the third response are not suitable to be provided in parallel on different output channels of the device.

[0026] In some embodiments, the method further includes: detecting that the digital assistant is operating in a hands-free or eyes-free mode; and upon detecting that the digital assistant is operating in the hands-free or eyes-free mode, determining that the second response and the third response are not suitable to be provided in parallel on different output channels.

[0027] In some embodiments, the method further includes: detecting that the digital assistant is operating in a hands-free or eyes-free mode; and providing the second response and the third response as respective speech outputs over an audio output channel of the device.

[0028] In some embodiments, the first information provision process further includes: processing the first speech input to identify an information request expressed in the first speech input; identifying a plurality of information items to fulfill the information request; and preparing a series of discrete sub-responses to the first speech input, the series of discrete sub-responses including at least the first response and the second response each describing respective one or more of the plurality of in-

formation items, and the series of sub-responses together constituting a complete response to the first speech input.

[0029] In some embodiments, the method further includes: determining a respective default time for delivering each of the series of sub-responses to the first speech input. In some embodiments, the method further includes dynamically overriding the respective default time for delivering the second response based on the determined relative urgency between the second response and the third response. In some embodiments, the method further includes: determining a respective default time for delivering the third response to the user; and dynamically overriding the respective default time for delivering the third response based on the determined relative urgency between the second response and the third response.

[0030] In some embodiments, the information request is a navigation request, the second response is a navigation instruction along a predetermined route, and wherein the first information provision process further includes determining a respective default time for providing the navigation instruction during vehicle navigation. In some embodiments, determining the respective default time for providing the navigation instruction further includes: determining the respective default time based on a predetermined proximity between a current location of the user and a respective waypoint associated with the second response along the predetermined route. In some embodiments, the method further includes: determining a respective default time for delivering the third response; and dynamically overriding at least one of the respective default delivering times for delivering the second and the third responses. In some embodiments, the respective default time for delivering the third response is a time immediately after the third response become available. In some embodiments, the respective default time for delivering the third response is a time immediately after the occurrence of a predetermined trigger event for the third response. In some embodiments, the predetermined trigger event is the arrival of a predetermined trigger time based on a system clock. In some embodiments, the predetermined trigger event is the arrival of a predetermined trigger message at the device.

[0031] In some embodiments, determining the relative urgency between the second response and the third response further includes: determining a present context associated with the user; and determining the relative urgency between the second response and the third response based the present context associated with the user. In some embodiments, the present context associated with the user is formed by one or more of: a current location of the user, a current speed of the user, a current travel direction of the user, a current time, one or more predetermined user preferences, a location associated with the second response, a location associated with the third response, a time period required to deliver the second response to the user, a time period required to deliver the third response to the user, a default time for delivering the second response to the user, a default time for delivering the third response to the user, a degree of conflict between the second response and the third response, content of the second response, content of the third response, a time window in which the second response will likely remain relevant, a time window in which the third response will likely remain relevant, a likely consequence for delivering the second response before the third response, and a likely consequence for delivering the third response before the second response.

[0032] In some embodiments, determining the relative urgency between the second response and the third response further includes determining whether the second response is associated with a location within a predetermined distance from a current location of the user. In some embodiments, the predetermined distance is based on a current speed of the user.

[0033] In some embodiments, determining the relative urgency between the second response and the third response further includes determining whether the user is likely to pass by a location associated with the second response within a predetermined time window from a current time.

[0034] In some embodiments, the predetermined time window is based on a respective time period required to deliver the third response to the user.

[0035] In some embodiments, determining the relative urgency between the second response and the third response further includes: upon determining that the user is likely to pass by the location associated with the second response within the predetermined time window from the current time, assigning a higher level of urgency to the second response than the third response.

[0036] In some embodiments, determining the relative urgency between the second response and the third response further includes: upon determining that the user is unlikely to pass by the location associated with the second response within the predetermined time window from the current time, assigning a higher level of urgency to the third response than the second response.

[0037] In some embodiments, determining the present context associated with the user further includes: monitoring a current location, a current direction, and a current speed of the user; and determining the present context based on the monitoring.

[0038] In some embodiments, the second response is provision of an information item in response to the first speech input, and the third response is an alert item generated for a previously established reminder or arrival of a push notification, and wherein the method further includes: based on the present context, determining whether delivery of the second response at a respective default delivery time of the second response is likely to adversely affect a utility of the reminder or push-notification to the user. In some embodiments, the method further includes: upon determining that delivery of the second response at the respective default delivery time is likely to affect the utility of the reminder or notification, assigning a lower

relative urgency to the second response than the third response; and delaying delivery of the second response until after delivery of the reminder or push-notification to the user. In some embodiments, the method further includes: upon determining that delivery of the second response at the respective default delivery time is unlikely to affect the utility of the reminder or notification, assigning a higher relative urgency to the second response than the third response; and delaying delivery of the alert item for the reminder or push-notification until after delivery of the second response to the user.

[0039] In some embodiments, detecting the event operable to initiate the second information provision process further includes receiving a second speech input prior to provision of any response to the first speech input, and receipt of the second speech input is operable to initiate the second information process.

[0040] In some embodiments, detecting the event operable to initiate the second information provision process further includes receiving a second speech input after provision of at least one of a series of responses to the first speech input and prior to provision of all of the series of responses to the user.

[0041] In some embodiments, the method further includes: detecting whether there is a timing conflict between delivery of the second response and the third response based on respective default delivery times of the second response and the third response; determining the relative urgency between the second response and the third response upon detection of the timing conflict; and overriding the respective default delivery time of one of the second and third responses based on the determined relative urgency.

[0042] In some embodiments, the respective default delivery time for the second response is a projected delivery time for the second response as if the second response were the only response waiting to be delivered to the user.

[0043] In some embodiments, the respective default delivery time for the third response is a projected delivery time for the third response as if the third response were the only response waiting to be delivered to the user.

[0044] In some embodiments, detecting whether there is the timing conflict is further based on respective durations of the second response and the third response when delivered to the user.

[0045] In some embodiments, a method includes features of any combination of the methods described above. In some embodiments, a non-transitory computer readable medium has instructions stored thereon, the instructions, when executed by one or more processors, cause the processors to perform any of the methods described above. In some embodiments, a system including one or more processors; and memory having instructions stored thereon, the instructions, when executed by the one or more processors, cause the processors to perform any of the methods described above.

[0046] In some examples a method of operating a digital assistant, comprises: at a device having one or more processors and memory: receiving a navigation request from a user; initiating a first information provision process in response to the navigation request, the first information provision process comprising preparing at least a first navigation instruction and a second navigation instruction; delivering the first navigation instruction to the user at a respective default delivery time associated with the first navigation instruction; after or concurrent with the delivery of the first navigation instruction, detecting an event operable to initiate a second information provision process; initiating the second information provision process in response to detecting the event, the second information process comprising preparing a respective output to be delivered to the user regarding the event; determining a relative urgency between the second navigation instruction and the output regarding the event; and providing the second navigation instruction and the output regarding the event in an order based on the determined relative urgency.

[0047] In some embodiments, the method further includes: determining respective default delivery times for the second navigation instruction and the output regarding the event; and determining whether there is a timing conflict between deliveries of the second navigation instruction and the output regarding the event according to their respective default delivery times.

[0048] In some embodiments, the method further includes: determining the relative urgency between the second navigation instruction and the output regarding the event upon detecting the timing conflict.

[0049] In some embodiments, the method further includes: overriding at least one of the respective default delivery times of the second navigation instruction and the output regarding the event based on the determined relative urgency.

[0050] In some embodiments, detecting the event further includes receiving an information request from the user, wherein the information request does not modify the directions request.

[0051] In some embodiments, the output regarding the event comprises at least a speech output containing information retrieved in response to the information request.

[0052] In some embodiments, detecting the event further comprises detecting occurrence of a trigger event for a previously established reminder. In some embodiments, the output regarding the event comprises at least an alert item providing content of the previously established reminder. In some embodiments, the respective default time for delivering the output regarding the event is a reminder time specified in a previously established reminder.

[0053] In some embodiments, detecting the event further includes: detecting arrival of a push-notification from an application or process not currently controlled by the digital assistant; and intercepting the push-notification before the push-notification is presented to the user. In

some embodiments, the output regarding the event is a speech output prepared by the digital assistant regarding the arrival of the push-notification. In some embodiments, the respective default time for delivering the output regarding the event is immediately after the arrival of the push-notification.

[0054] In some embodiments, providing the second navigation instruction and the output regarding the event in an order based on the determined relative urgency further includes: determining that the second navigation instruction has a higher relative urgency than the output regarding the event; and delivering the second navigation instruction before the output regarding the event.

[0055] In some embodiments, providing the second navigation instruction and the output regarding the event in an order based on the determined relative urgency further includes: determining that the second navigation instruction has a lower relative urgency than the output regarding the event; and delivering the second navigation instruction after the output regarding the event.

[0056] In some embodiments, the respective default time for providing the second navigation instruction is based on a predetermined proximity between a current location of the user and a respective waypoint associated with the second navigation instruction.

[0057] In some embodiments, determining the relative urgency between the second navigation instruction and the output regarding the event further comprises: determining a present context associated with the user; and determining the relative urgency between the second response and the third response based the present context associated with the user.

[0058] In some embodiments, determining the relative urgency between the second navigation instruction and the output regarding the event further includes: determining whether the second navigation instruction is associated with a waypoint within a predetermined distance from a current location of the user.

[0059] In some embodiments, determining the relative urgency between the second navigation instruction and the output regarding the event further includes: determining whether the user is likely to pass by a waypoint associated with the second navigation instruction within a predetermined time window from a current time. In some embodiments, the predetermined time window is based on a respective time buffer required to deliver the third response to the user.

[0060] In some embodiments, determining the relative urgency between the second navigation instruction and the output regarding the event further includes: upon determining that the user is likely to pass by the waypoint associated with the second navigation instruction within the predetermined time window from the current time, assigning a higher level of urgency to the second navigation instruction than the output regarding the event.

[0061] In some embodiments, determining the relative urgency between the second navigation instruction and the output regarding the event further includes: upon determining that the user is unlikely to pass by the waypoint associated with the second navigation instruction within the predetermined time window from the current time, assigning a higher level of urgency to the second navigation instruction than the output regarding the event.

[0062] In some embodiments, determining the present context associated with the user further includes: monitoring a current location, a current direction, and a current speed of the user; and determining the present context based on the monitoring.

[0063] In some embodiments, the output regarding the event is an alert item generated for a previously established reminder or arrival of a third-party notification, and determining the relative urgency further includes: based on the present context, determining whether delivery of the second navigation instruction at the respective default delivery time of the second navigation instruction is likely to adversely affect a utility of the reminder or notification to the user.

[0064] In some embodiments, the method further includes: upon determining that delivery of the second navigation instruction at the respective default delivery time of the second navigation instruction is likely to affect the utility of the reminder or notification, assigning a lower relative urgency to the second navigation instruction than the alert item; and delaying delivery of the second navigation instruction until after delivery of the alert item for the reminder or notification to the user.

[0065] In some embodiments, the method further includes: upon determining that delivery of the second navigation instruction at the respective default delivery time of the second navigation instruction is unlikely to affect the utility of the reminder or notification, assigning a higher relative urgency to the second navigation instruction than the alert item for the reminder or notification; and delaying delivery of the alert item for the reminder or notification until after delivery of the second navigation instruction to the user.

[0066] In some embodiments, the utility of the reminder or notification is unlikely to be affected during a period between a reminder time that is specified in the previously established reminder and a threshold amount of preparation time needed before an event time specified in the previously established reminder.

[0067] In accordance with some examples an electronic device includes one or more processors, memory, and one or more programs; the one or more programs are stored in the memory and configured to be executed by the one or more processors and the one or more programs include instructions for performing the operations of any of the methods described above. In accordance with some examples , a computer readable storage medium has stored therein instructions, which, when executed by an electronic device, cause the device to perform the operations of any of the methods described above. In accordance with some examples , an electronic device includes means for performing the operations of any of the methods described above. In accordance with

some examples an information processing apparatus, for use in an electronic device includes means for performing the operations of any of the methods described above.

[0068] In accordance with some examples , an electronic device includes a speech receiving unit configured to receive a first speech input from a user; a first information provision unit configured to initiate a first information provision process in response to receipt of the first speech input, the first information provision process comprising preparing at least a first response and a second response to the first speech input; a first response provision unit configured to provide the first response to the user; a detection unit configured to, after or concurrent with the provision of the first response to the user, but before provision of the second response to the user, detect an event operable to initiate a second information provision process; a second information provision unit configured to initiate the second information provision process in response to detecting the event, the second information provision process comprising preparing at least a third response to the event; a urgency determining unit configured to determine a relative urgency between the second response and the third response; and a second response provision unit configured to provide one of the second response and the third response to the user in an order based on the determined relative urgency.

[0069] In accordance with some examples , an electronic device includes a request receiving unit configured to receive a navigation request from a user; a first information provision unit configured to initiate a first information provision process in response to the navigation request, the first information provision process comprising preparing at least a first navigation instruction and a second navigation instruction; a delivering unit configured to deliver the first navigation instruction to the user at a respective default delivery time associated with the first navigation instruction; a detection unit configured to, after or concurrent with the delivery of the first navigation instruction, detect an event operable to initiate a second information provision process; a second information provision unit configured to initiate the second information provision process in response to detecting the event, the second information process comprising preparing a respective output to be delivered to the user regarding the event; an urgency determining unit configured to determine a relative urgency between the second navigation instruction and the output regarding the event; and an instruction provision unit configured to provide the second navigation instruction and the output regarding the event in an order based on the determined relative urgency.

[0070] The details of one or more embodiments of the subject matter described in this specification are set forth in the accompanying drawings and the description below. Other features, aspects, and advantages of the subject matter will become apparent from the description, the drawings, and the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[0071]

FIG. 1 is a block diagram illustrating an environment in which a digital assistant operates in accordance with some embodiments.

FIG. 2 is a block diagram illustrating a digital assistant client system in accordance with some embodiments.

FIG. 3A is a block diagram illustrating a digital assistant system or a server portion thereof in accordance with some embodiments.

FIG. 3B is a block diagram illustrating functions of the digital assistant shown in FIG. 3A in accordance with some embodiments.

FIG. 3C is a diagram of a portion of an ontology in accordance with some embodiments.

FIG. 4 is a flow chart for an exemplary process for intelligently handling a user-initiated or system-initiated interruption of an existing task in accordance with some embodiments.

FIG. 5A is a diagram illustrating an example scenario in which intelligent handling of user-initiated and system-initiated interruptions while the digital assistant is delivering navigation instructions during vehicle navigation in accordance with some embodiments.

FIG. 5B is a flow chart for an exemplary process for intelligently handling a user-initiated or system-initiated interruption while the digital assistant is delivering navigation instructions during navigation in accordance with some embodiments.

FIG. 6 illustrates a functional block diagram of an electronic device in accordance with some embodiments.

FIG. 7 illustrates a functional block diagram of an electronic device in accordance with some embodiments.

[0072] Like reference numerals refer to corresponding parts throughout the drawings.

DESCRIPTION OF EMBODIMENTS

[0073] FIG. 1 is a block diagram of an operating environment 100 of a digital assistant according to some embodiments. The terms "digital assistant," "virtual assistant," "intelligent automated assistant," or "automatic digital assistant," refer to any information processing system

that interprets natural language input in spoken and/or textual form to deduce user intent, and performs actions based on the deduced user intent. For example, to act on a deduced user intent, the system can perform one or more of the following: identifying a task flow with steps and parameters designed to accomplish the deduced user intent, inputting specific requirements from the deduced user intent into the task flow; executing the task flow by invoking programs, methods, services, APIs, or the like; and generating output responses to the user in an audible (*e.g.* speech) and/or visual form.

[0074] Specifically, a digital assistant is capable of accepting a user request at least partially in the form of a natural language command, request, statement, narrative, and/or inquiry. Typically, the user request seeks either an informational answer or performance of a task by the digital assistant. A satisfactory response to the user request is either provision of the requested informational answer, performance of the requested task, or a combination of the two. For example, a user may ask the digital assistant a question, such as "Where am I right now?" Based on the user's current location, the digital assistant may answer, "You are in Central Park near the west gate." The user may also request the performance of a task, for example, "Please invite my friends to my girlfriend's birthday party next week." In response, the digital assistant may acknowledge the request by saying "Yes, right away," and then send a suitable calendar invite on behalf of the user to each of the user' friends listed in the user's electronic address book. During performance of a requested task, the digital assistant sometimes interacts with the user in a continuous dialogue involving multiple exchanges of information over an extended period of time. There are numerous other ways of interacting with a digital assistant to request information or performance of various tasks. In addition to providing verbal responses and taking programmed actions, the digital assistant also provides responses in other visual or audio forms, *e.g.,* as text, alerts, music, videos, animations, etc.

[0075] An example of a digital assistant is described in Applicant's U.S. Utility Application Serial No. 12/987,982 for "Intelligent Automated Assistant," filed January 10, 2011.

[0076] As shown in FIG. 1, in some embodiments, a digital assistant is implemented according to a client-server model. The digital assistant includes a client-side portion 102a, 102b (hereafter "DA client 102") executed on a user device 104a, 104b, and a server-side portion 106 (hereafter "DA server 106") executed on a server system 108. The DA client 102 communicates with the DA server 106 through one or more networks 110. The DA client 102 provides client-side functionalities such as user-facing input and output processing and communications with the DA server 106. The DA server 106 provides server-side functionalities for any number of DA clients 102 each residing on a respective user device 104.

[0077] In some embodiments, the DA server 106 includes a client-facing I/O interface 112, one or more processing modules 114, data and models 116, and an I/O interface to external services 118. The client-facing I/O interface facilitates the client-facing input and output processing for the digital assistant server 106. The one or more processing modules 114 utilize the data and models 116 to determine the user's intent based on natural language input and perform task execution based on deduced user intent. In some embodiments, the DA server 106 communicates with external services 120 through the network(s) 110 for task completion or information acquisition. The I/O interface to external services 118 facilitates such communications.

[0078] Examples of the user device 104 include, but are not limited to, a handheld computer, a personal digital assistant (PDA), a tablet computer, a laptop computer, a desktop computer, a cellular telephone, a smart phone, an enhanced general packet radio service (EGPRS) mobile phone, a media player, a navigation device, a game console, a television, a remote control, or a combination of any two or more of these data processing devices or other data processing devices. More details on the user device 104 are provided in reference to an exemplary user device 104 shown in FIG. 2.

[0079] Examples of the communication network(s) 110 include local area networks ("LAN") and wide area networks ("WAN"), *e.g.,* the Internet. The communication network(s) 110 may be implemented using any known network protocol, including various wired or wireless protocols, such as *e.g.,* Ethernet, Universal Serial Bus (USB), FIREWIRE, Global System for Mobile Communications (GSM), Enhanced Data GSM Environment (EDGE), code division multiple access (CDMA), time division multiple access (TDMA), Bluetooth, Wi-Fi, voice over Internet Protocol (VoIP), Wi-MAX, or any other suitable communication protocol.

[0080] The server system 108 is implemented on one or more standalone data processing apparatus or a distributed network of computers. In some embodiments, the server system 108 also employs various virtual devices and/or services of third party service providers (*e.g.,* third-party cloud service providers) to provide the underlying computing resources and/or infrastructure resources of the server system 108.

[0081] Although the digital assistant shown in FIG. 1 includes both a client-side portion (*e.g.,* the DA client 102) and a server-side portion (*e.g.,* the DA server 106), in some embodiments, the functions of a digital assistant is implemented as a standalone application installed on a user device. In addition, the divisions of functionalities between the client and server portions of the digital assistant can vary in different embodiments. For example, in some embodiments, the DA client is a thin-client that provides only user-facing input and output processing functions, and delegates all other functionalities of the digital assistant to a backend server.

[0082] FIG. 2 is a block diagram of a user-device 104 in accordance with some embodiments. The user device 104 includes a memory interface 202, one or more proc-

essors 204, and a peripherals interface 206. The various components in the user device 104 are coupled by one or more communication buses or signal lines. The user device 104 includes various sensors, subsystems, and peripheral devices that are coupled to the peripherals interface 206. The sensors, subsystems, and peripheral devices gather information and/or facilitate various functionalities of the user device 104.

[0083] For example, a motion sensor 210, a light sensor 212, and a proximity sensor 214 are coupled to the peripherals interface 206 to facilitate orientation, light, and proximity sensing functions. One or more other sensors 216, such as a positioning system (*e.g.,* GPS receiver), a temperature sensor, a biometric sensor, a gyro, a compass, an accelerometer, and the like, are also connected to the peripherals interface 206, to facilitate related functionalities.

[0084] In some embodiments, a camera subsystem 220 and an optical sensor 222 are utilized to facilitate camera functions, such as taking photographs and recording video clips. Communication functions are facilitated through one or more wired and/or wireless communication subsystems 224, which can include various communication ports, radio frequency receivers and transmitters, and/or optical (*e.g.,* infrared) receivers and transmitters. An audio subsystem 226 is coupled to speakers 228 and a microphone 230 to facilitate voice-enabled functions, such as voice recognition, voice replication, digital recording, and telephony functions.

[0085] In some embodiments, an I/O subsystem 240 is also coupled to the peripheral interface 206. The I/O subsystem 240 includes a touch screen controller 242 and/or other input controller(s) 244. The touch-screen controller 242 is coupled to a touch screen 246. The touch screen 246 and the touch screen controller 242 can, for example, detect contact and movement or break thereof using any of a plurality of touch sensitivity technologies, such as capacitive, resistive, infrared, surface acoustic wave technologies, proximity sensor arrays, and the like. The other input controller(s) 244 can be coupled to other input/control devices 248, such as one or more buttons, rocker switches, thumb-wheel, infrared port, USB port, and/or a pointer device such as a stylus.

[0086] In some embodiments, the memory interface 202 is coupled to memory 250. The memory 250 can include high-speed random access memory and/or non-volatile memory, such as one or more magnetic disk storage devices, one or more optical storage devices, and/or flash memory (*e.g.*, NAND, NOR).

[0087] In some embodiments, the memory 250 stores an operating system 252, a communication module 254, a user interface module 256, a sensor processing module 258, a phone module 260, and applications 262. The operating system 252 includes instructions for handling basic system services and for performing hardware dependent tasks. The communication module 254 facilitates communicating with one or more additional devices, one or more computers and/or one or more servers. The user interface module 256 facilitates graphic user interface processing and output processing using other output channels (*e.g.,* speakers). The sensor processing module 258 facilitates sensor-related processing and functions. The phone module 260 facilitates phone-related processes and functions. The application module 262 facilitates various functionalities of user applications, such as electronic-messaging, web browsing, media processing, Navigation, imaging and/or other processes and functions.

[0088] As described in this specification, the memory 250 also stores client-side digital assistant instructions (*e.g.,* in a digital assistant client module 264) and various user data 266 (*e.g.,* user-specific vocabulary data, preference data, and/or other data such as the user's electronic address book, to-do lists, shopping lists, etc.) to provide the client-side functionalities of the digital assistant.

[0089] In various embodiments, the digital assistant client module 264 is capable of accepting voice input (*e.g.,* speech input), text input, touch input, and/or gestural input through various user interfaces (*e.g.,* the I/O subsystem 240) of the user device 104. The digital assistant client module 264 is also capable of providing output in audio (*e.g.,* speech output), visual, and/or tactile forms. For example, output can be provided as voice, sound, alerts, text messages, menus, graphics, videos, animations, vibrations, and/or combinations of two or more of the above. During operation, the digital assistant client module 264 communicates with the digital assistant server using the communication subsystems 224.

[0090] In some embodiments, the digital assistant client module 264 utilizes the various sensors, subsystems and peripheral devices to gather additional information from the surrounding environment of the user device 104 to establish a context associated with a user, the current user interaction, and/or the current user input. In some embodiments, the digital assistant client module 264 provides the context information or a subset thereof with the user input to the digital assistant server to help deduce the user's intent. In some embodiments, the digital assistant also uses the context information to determine how to prepare and delivery outputs to the user.

[0091] In some embodiments, the context information that accompanies the user input includes sensor information, *e.g.,* lighting, ambient noise, ambient temperature, images or videos of the surrounding environment, etc. In some embodiments, the context information also includes the physical state of the device, *e.g.,* device orientation, device location, device temperature, power level, speed, acceleration, motion patterns, cellular signals strength, etc. In some embodiments, information related to the software state of the user device 104, *e.g.,* running processes, installed programs, past and present network activities, background services, error logs, resources usage, etc., of the user device 104 are provided to the digital assistant server as context information associated with a user input.

[0092] In some embodiments, the DA client module 264 selectively provides information (*e.g.,* user data 266) stored on the user device 104 in response to requests from the digital assistant server. In some embodiments, the digital assistant client module 264 also elicits additional input from the user via a natural language dialogue or other user interfaces upon request by the digital assistant server 106. The digital assistant client module 264 passes the additional input to the digital assistant server 106 to help the digital assistant server 106 in intent deduction and/or fulfillment of the user's intent expressed in the user request.

[0093] In various embodiments, the memory 250 includes additional instructions or fewer instructions. Furthermore, various functions of the user device 104 may be implemented in hardware and/or in firmware, including in one or more signal processing and/or application specific integrated circuits.

[0094] FIG. 3A is a block diagram of an example digital assistant system 300 in accordance with some embodiments. In some embodiments, the digital assistant system 300 is implemented on a standalone computer system. In some embodiments, the digital assistant system 300 is distributed across multiple computers. In some embodiments, some of the modules and functions of the digital assistant are divided into a server portion and a client portion, where the client portion resides on a user device (*e.g.,* the user device 104) and communicates with the server portion (*e.g.,* the server system 108) through one or more networks, *e.g.*, as shown in FIG. 1. In some embodiments, the digital assistant system 300 is an embodiment of the server system 108 (and/or the digital assistant server 106) shown in FIG. 1. It should be noted that the digital assistant system 300 is only one example of a digital assistant system, and that the digital assistant system 300 may have more or fewer components than shown, may combine two or more components, or may have a different configuration or arrangement of the components. The various components shown in FIG. 3A may be implemented in hardware, software instructions for execution by one or more processors, firmware, including one or more signal processing and/or application specific integrated circuits, or a combination of thereof.

[0095] The digital assistant system 300 includes memory 302, one or more processors 304, an input/output (I/O) interface 306, and a network communications interface 308. These components communicate with one another over one or more communication buses or signal lines 310.

[0096] In some embodiments, the memory 302 includes a non-transitory computer readable medium, such as high-speed random access memory and/or a non-volatile computer readable storage medium (*e.g.,* one or more magnetic disk storage devices, flash memory devices, or other non-volatile solid-state memory devices).

[0097] In some embodiments, the I/O interface 306 couples input/output devices 316 of the digital assistant system 300, such as displays, a keyboards, touch screens, and microphones, to the user interface module 322. The I/O interface 306, in conjunction with the user interface module 322, receive user inputs (*e.g.,* voice input, keyboard inputs, touch inputs, etc.) and process them accordingly. In some embodiments, *e.g.,* when the digital assistant is implemented on a standalone user device, the digital assistant system 300 includes any of the components and I/O and communication interfaces described with respect to the user device 104 in FIG. 2. In some embodiments, the digital assistant system 300 represents the server portion of a digital assistant implementation, and interacts with the user through a client-side portion residing on a user device (*e.g.,* the user device 104 shown in FIG. 2).

[0098] In some embodiments, the network communications interface 308 includes wired communication port(s) 312 and/or wireless transmission and reception circuitry 314. The wired communication port(s) receive and send communication signals via one or more wired interfaces, *e.g.,* Ethernet, Universal Serial Bus (USB), FIREWIRE, etc. The wireless circuitry 314 receives and sends RF signals and/or optical signals from/to communications networks and other communications devices. The wireless communications may use any of a plurality of communications standards, protocols and technologies, such as GSM, EDGE, CDMA, TDMA, Bluetooth, Wi-Fi, VoIP, Wi-MAX, or any other suitable communication protocol. The network communications interface 308 enables communication between the digital assistant system 300 with networks, such as the Internet, an intranet and/or a wireless network, such as a cellular telephone network, a wireless local area network (LAN) and/or a metropolitan area network (MAN), and other devices.

[0099] In some embodiments, memory 302, or the computer readable storage media of memory 302, stores programs, modules, instructions, and data structures including all or a subset of: an operating system 318, a communications module 320, a user interface module 322, one or more applications 324, and a digital assistant module 326. The one or more processors 304 execute these programs, modules, and instructions, and reads/writes from/to the data structures.

[0100] The operating system 318 (*e.g.,* Darwin, RTXC, LINUX, UNIX, OS X, WINDOWS, or an embedded operating system such as VxWorks) includes various software components and/or drivers for controlling and managing general system tasks (*e.g.,* memory management, storage device control, power management, etc.) and facilitates communications between various hardware, firmware, and software components.

[0101] The communications module 320 facilitates communications between the digital assistant system 300 with other devices over the network communications interface 308. For example, the communication module 320 may communicate with the communication interface 254 of the device 104 shown in FIG. 2. The communications module 320 also includes various components for

handling data received by the wireless circuitry 314 and/or wired communications port 312.

[0102] The user interface module 322 receives commands and/or inputs from a user via the I/O interface 306 (*e.g.,* from a keyboard, touch screen, pointing device, controller, and/or microphone), and generates user interface objects on a display. The user interface module 322 also prepares and delivers outputs (*e.g.,* speech, sound, animation, text, icons, vibrations, haptic feedback, and light, etc.) to the user via the I/O interface 306 (*e.g.,* through displays, audio channels, speakers, and touch-pads, etc.).

[0103] The applications 324 include programs and/or modules that are configured to be executed by the one or more processors 304. For example, if the digital assistant system is implemented on a standalone user device, the applications 324 may include user applications, such as games, a calendar application, a navigation application, or an email application. If the digital assistant system 300 is implemented on a server farm, the applications 324 may include resource management applications, diagnostic applications, or scheduling applications, for example.

[0104] The memory 302 also stores the digital assistant module (or the server portion of a digital assistant) 326. In some embodiments, the digital assistant module 326 includes the following sub-modules, or a subset or superset thereof: an input/output processing module 328, a speech-to-text (STT) processing module 330, a natural language processing module 332, a dialogue flow processing module 334, a task flow processing module 336, a service processing module 338, and an interruption handling module 340. Each of these modules has access to one or more of the following data and models of the digital assistant 326, or a subset or superset thereof: ontology 360, vocabulary index 344, user data 348, task flow models 354, service models 356, and priority parameters database 358.

[0105] In some embodiments, using the processing modules, data, and models implemented in the digital assistant module 326, the digital assistant performs at least some of the following: identifying a user's intent expressed in a natural language input received from the user; actively eliciting and obtaining information needed to fully deduce the user's intent (*e.g.,* by disambiguating words, names, intentions, etc.); determining the task flow for fulfilling the deduced intent; and executing the task flow to fulfill the deduced intent. In this specification, more details regarding the interruption handling module and its use of the priority parameters are provided later in FIGS. 4-5B and accompanying descriptions.

[0106] In some embodiments, as shown in FIG. 3B, the I/O processing module 328 interacts with the user through the I/O devices 316 in FIG. 3A or with a user device (*e.g.,* a user device 104 in FIG. 1) through the network communications interface 308 in FIG. 3A to obtain user input (*e.g.,* a speech input) and to provide responses (*e.g.,* as speech outputs) to the user input. The I/O processing module 328 optionally obtains context information associated with the user input from the user device, along with or shortly after the receipt of the user input. The context information includes user-specific data, vocabulary, and/or preferences relevant to the user input. In some embodiments, the context information also includes software and hardware states of the device (e.g., the user device 104 in FIG. 1) at the time the user request is received, and/or information related to the surrounding environment of the user at the time that the user request was received. In some embodiments, the I/O processing module 328 also sends follow-up questions to, and receives answers from, the user regarding the user request. When a user request is received by the I/O processing module 328 and the user request contains a speech input, the I/O processing module 328 forwards the speech input to the speech-to-text (STT) processing module 330 for speech-to-text conversions.

[0107] The speech-to-text processing module 330 receives speech input (*e.g.,* a user utterance captured in a voice recording) through the I/O processing module 328. In some embodiments, the speech-to-text processing module 330 uses various acoustic and language models to recognize the speech input as a sequence of phonemes, and ultimately, a sequence of words or tokens written in one or more languages. The speech-to-text processing module 330 can be implemented using any suitable speech recognition techniques, acoustic models, and language models, such as Hidden Markov Models, Dynamic Time Warping (DTW)-based speech recognition, and other statistical and/or analytical techniques. In some embodiments, the speech-to-text processing can be performed at least partially by a third party service or on the user's device. Once the speech-to-text processing module 330 obtains the result of the speech-to-text processing, *e.g.,* a sequence of words or tokens, it passes the result to the natural language processing module 332 for intent deduction.

[0108] More details on the speech-to-text processing are described in U.S. Utility Application Serial No. 13/236,942 for "Consolidating Speech Recognition Results," filed on September 20, 2011.

[0109] The natural language processing module 332 ("natural language processor") of the digital assistant takes the sequence of words or tokens ("token sequence") generated by the speech-to-text processing module 330, and attempts to associate the token sequence with one or more "actionable intents" recognized by the digital assistant. An "actionable intent" represents a task that can be performed by the digital assistant, and has an associated task flow implemented in the task flow models 354. The associated task flow is a series of programmed actions and steps that the digital assistant takes in order to perform the task. The scope of a digital assistant's capabilities is dependent on the number and variety of task flows that have been implemented and stored in the task flow models 354, or in other words, on the number and variety of "actionable intents" that the

digital assistant recognizes. The effectiveness of the digital assistant, however, is also dependent on the assistant's ability to deduce the correct "actionable intent(s)" from the user request expressed in natural language.

[0110] In some embodiments, in addition to the sequence of words or tokens obtained from the speech-to-text processing module 330, the natural language processor 332 also receives context information associated with the user request, *e.g.,* from the I/O processing module 328. The natural language processor 332 optionally uses the context information to clarify, supplement, and/or further define the information contained in the token sequence received from the speech-to-text processing module 330. The context information includes, for example, user preferences, hardware and/or software states of the user device, sensor information collected before, during, or shortly after the user request, prior interactions (*e.g.,* dialogue) between the digital assistant and the user, and the like. As described in this specification, context information is dynamic, and can change with time, location, content of the dialogue, and other factors.

[0111] In some embodiments, the natural language processing is based on ontology 360. The ontology 360 is a hierarchical structure containing many nodes, each node representing either an "actionable intent" or a "property" relevant to one or more of the "actionable intents" or other "properties". As noted above, an "actionable intent" represents a task that the digital assistant is capable of performing, *i.e.,* it is "actionable" or can be acted on. A "property" represents a parameter associated with an actionable intent or a sub-aspect of another property. A linkage between an actionable intent node and a property node in the ontology 360 defines how a parameter represented by the property node pertains to the task represented by the actionable intent node.

[0112] In some embodiments, the ontology 360 is made up of actionable intent nodes and property nodes. Within the ontology 360, each actionable intent node is linked to one or more property nodes either directly or through one or more intermediate property nodes. Similarly, each property node is linked to one or more actionable intent nodes either directly or through one or more intermediate property nodes. For example, as shown in FIG. 3C, the ontology 360 may include a "restaurant reservation" node (*i.e.,* an actionable intent node). Property nodes "restaurant," "date/time" (for the reservation), and "party size" are each directly linked to the actionable intent node (*i.e.,* the "restaurant reservation" node). In addition, property nodes "cuisine," "price range," "phone number," and "location" are sub-nodes of the property node "restaurant," and are each linked to the "restaurant reservation" node (i.e., the actionable intent node) through the intermediate property node "restaurant." For another example, as shown in FIG. 3C, the ontology 360 may also include a "set reminder" node (i.e., another actionable intent node). Property nodes "date/time" (for the setting the reminder) and "subject" (for the reminder) are each linked to the "set reminder" node. Since the property "date/time" is relevant to both the task of making a restaurant reservation and the task of setting a reminder, the property node "date/time" is linked to both the "restaurant reservation" node and the "set reminder" node in the ontology 360.

[0113] An actionable intent node, along with its linked concept nodes, may be described as a "domain." In the present discussion, each domain is associated with a respective actionable intent, and refers to the group of nodes (and the relationships therebetween) associated with the particular actionable intent. For example, the ontology 360 shown in FIG. 3C includes an example of a restaurant reservation domain 362 and an example of a reminder domain 364 within the ontology 360. The restaurant reservation domain includes the actionable intent node "restaurant reservation," property nodes "restaurant," "date/time," and "party size," and sub-property nodes "cuisine," "price range," "phone number," and "location." The reminder domain 364 includes the actionable intent node "set reminder," and property nodes "subject" and "date/time." In some embodiments, the ontology 360 is made up of many domains. Each domain may share one or more property nodes with one or more other domains. For example, the "date/time" property node may be associated with many different domains (*e.g.,* a scheduling domain, a travel reservation domain, a movie ticket domain, etc.), in addition to the restaurant reservation domain 362 and the reminder domain 364.

[0114] While FIG. 3C illustrates two example domains within the ontology 360, other domains (or actionable intents) include, for example, "initiate a phone call," "find directions," "schedule a meeting," "send a message," and "provide an answer to a question," "read a list", "providing navigation instructions," "provide instructions for a task" and so on. A "send a message" domain is associated with a "send a message" actionable intent node, and may further include property nodes such as "recipient(s)", "message type", and "message body." The property node "recipient" may be further defined, for example, by the sub-property nodes such as "recipient name" and "message address."

[0115] In some embodiments, the ontology 360 includes all the domains (and hence actionable intents) that the digital assistant is capable of understanding and acting upon. In some embodiments, the ontology 360 may be modified, such as by adding or removing entire domains or nodes, or by modifying relationships between the nodes within the ontology 360.

[0116] In some embodiments, nodes associated with multiple related actionable intents may be clustered under a "super domain" in the ontology 360. For example, a "travel" super-domain may include a cluster of property nodes and actionable intent nodes related to travels. The actionable intent nodes related to travels may include "airline reservation," "hotel reservation," "car rental," "get directions," "find points of interest," and so on. The actionable intent nodes under the same super domain (e.g., the "travels" super domain) may have many property

nodes in common. For example, the actionable intent nodes for "airline reservation," "hotel reservation," "car rental," "get directions," "find points of interest" may share one or more of the property nodes "start location," "destination," "departure date/time," "arrival date/time," and "party size."

[0117] In some embodiments, each node in the ontology 360 is associated with a set of words and/or phrases that are relevant to the property or actionable intent represented by the node. The respective set of words and/or phrases associated with each node is the so-called "vocabulary" associated with the node. The respective set of words and/or phrases associated with each node can be stored in the vocabulary index 344 in association with the property or actionable intent represented by the node. For example, returning to FIG. 3B, the vocabulary associated with the node for the property of "restaurant" may include words such as "food," "drinks," "cuisine," "hungry," "eat," "pizza," "fast food," "meal," and so on. For another example, the vocabulary associated with the node for the actionable intent of "initiate a phone call" may include words and phrases such as "call," "phone," "dial," "ring," "call this number," "make a call to," and so on. The vocabulary index 344 optionally includes words and phrases in different languages.

[0118] The natural language processor 332 receives the token sequence (e.g., a text string) from the speech-to-text processing module 330, and determines what nodes are implicated by the words in the token sequence. In some embodiments, if a word or phrase in the token sequence is found to be associated with one or more nodes in the ontology 360 (via the vocabulary index 344), the word or phrase will "trigger" or "activate" those nodes. Based on the quantity and/or relative importance of the activated nodes, the natural language processor 332 will select one of the actionable intents as the task that the user intended the digital assistant to perform. In some embodiments, the domain that has the most "triggered" nodes is selected. In some embodiments, the domain having the highest confidence value (*e.g.,* based on the relative importance of its various triggered nodes) is selected. In some embodiments, the domain is selected based on a combination of the number and the importance of the triggered nodes. In some embodiments, additional factors are considered in selecting the node as well, such as whether the digital assistant has previously correctly interpreted a similar request from a user.

[0119] In some embodiments, the digital assistant also stores names of specific entities in the vocabulary index 344, so that when one of these names is detected in the user request, the natural language processor 332 will be able to recognize that the name refers to a specific instance of a property or sub-property in the ontology. In some embodiments, the names of specific entities are names of businesses, restaurants, people, movies, and the like. In some embodiments, the digital assistant searches and identifies specific entity names from other data sources, such as the user's address book, a movies database, a musicians database, and/or a restaurant database. In some embodiments, when the natural language processor 332 identifies that a word in the token sequence is a name of a specific entity (such as a name in the user's address book), that word is given additional significance in selecting the actionable intent within the ontology for the user request.

[0120] For example, when the words "Mr. Santo" are recognized from the user request and the last name "Santo" is found in the vocabulary index 344 as one of the contacts in the user's contact list, then it is likely that the user request corresponds to a "send a message" or "initiate a phone call" domain. For another example, when the words "ABC Cafe" are found in the user request, and the term "ABC Cafe" is found in the vocabulary index 344 as the name of a particular restaurant in the user's city, then it is likely that the user request corresponds to a "restaurant reservation" domain.

[0121] User data 348 includes user-specific information, such as user-specific vocabulary, user preferences, user address, user's default and secondary languages, user's contact list, and other short-term or long-term information for each user. In some embodiments, the natural language processor 332 uses the user-specific information to supplement the information contained in the user input to further define the user intent. For example, for a user request "invite my friends to my birthday party," the natural language processor 332 is able to access user data 348 to determine who the "friends" are and when and where the "birthday party" would be held, rather than requiring the user to provide such information explicitly in his/her request.

[0122] Other details of searching an ontology based on a token string is described in U.S. Utility Application Serial No. 12/341,743 for "Method and Apparatus for Searching Using An Active Ontology," filed December 22, 2008.

[0123] In some embodiments, once the natural language processor 332 identifies an actionable intent (or domain) based on the user request, the natural language processor 332 generates a structured query to represent the identified actionable intent. In some embodiments, the structured query includes parameters for one or more nodes within the domain for the actionable intent, and at least some of the parameters are populated with the specific information and requirements specified in the user request. For example, the user may say "Make me a dinner reservation at a sushi place at 7." In this case, the natural language processor 332 may be able to correctly identify the actionable intent to be "restaurant reservation" based on the user input. According to the ontology, a structured query for a "restaurant reservation" domain may include parameters such as {Cuisine}, {Time}, {Date}, {Party Size}, and the like. In some embodiments, based on the information contained in the user's utterance, the natural language processor 332 generates a partial structured query for the restaurant reservation domain, where the partial structured query includes the pa-

rameters {Cuisine= "Sushi"} and {Time = "7pm"}. However, in this example, the user's utterance contains insufficient information to complete the structured query associated with the domain. Therefore, other necessary parameters such as {Party Size} and {Date} are not specified in the structured query based on the information currently available. In some embodiments, the natural language processor 332 populates some parameters of the structured query with received context information. For example, in some embodiments, if the user requested a sushi restaurant "near me," the natural language processor 332 populates a {location} parameter in the structured query with GPS coordinates from the user device 104.

[0124] In some embodiments, the natural language processor 332 passes the structured query (including any completed parameters) to the task flow processing module 336 ("task flow processor"). The task flow processor 336 is configured to receive the structured query from the natural language processor 332, complete the structured query, if necessary, and perform the actions required to "complete" the user's ultimate request. In some embodiments, the various procedures necessary to complete these tasks are provided in task flow models 354. In some embodiments, the task flow models include procedures for obtaining additional information from the user, and task flows for performing actions associated with the actionable intent.

[0125] As described above, in order to complete a structured query, the task flow processor 336 may need to initiate additional dialogue with the user in order to obtain additional information, and/or disambiguate potentially ambiguous utterances. When such interactions are necessary, the task flow processor 336 invokes the dialogue processing module 334 ("dialogue processor 334") to engage in a dialogue with the user. In some embodiments, the dialogue processor 334 determines how (and/or when) to ask the user for the additional information, and receives and processes the user responses. The questions are provided to and answers are received from the users through the I/O processing module 328. In some embodiments, the dialogue processor 334 presents dialogue output to the user via audio and/or visual output, and receives input from the user via spoken or physical (*e.g.,* clicking) responses. Continuing with the example above, when the task flow processor 336 invokes the dialogue flow processor 334 to determine the "party size" and "date" information for the structured query associated with the domain "restaurant reservation," the dialogue flow processor 334 generates questions such as "For how many people?" and "On which day?" to pass to the user. Once answers are received from the user, the dialogue flow processor 334 can then populate the structured query with the missing information, or pass the information to the task flow processor 336 to complete the missing information from the structured query.

[0126] In some cases, the task flow processor 336 may receive a structured query that has one or more ambiguous properties. For example, a structured query for the "send a message" domain may indicate that the intended recipient is "Bob," and the user may have multiple contacts named "Bob." The task flow processor 336 will request that the dialogue processor 334 disambiguate this property of the structured query. In turn, the dialogue processor 334 may ask the user "Which Bob?", and display (or read) a list of contacts named "Bob" from which the user may choose.

[0127] Once the task flow processor 336 has completed the structured query for an actionable intent, the task flow processor 336 proceeds to perform the ultimate task associated with the actionable intent. Accordingly, the task flow processor 336 executes the steps and instructions in the task flow model according to the specific parameters contained in the structured query. For example, the task flow model for the actionable intent of "restaurant reservation" may include steps and instructions for contacting a restaurant and actually requesting a reservation for a particular party size at a particular time. For example, using a structured query such as: {restaurant reservation, restaurant = ABC Café, date = 3/12/2012, time = 7pm, party size = 5}, the task flow processor 336 may perform the steps of: (1) logging onto a server of the ABC Café or a restaurant reservation system such as OPENTABLE®, (2) entering the date, time, and party size information in a form on the website, (3) submitting the form, and (4) making a calendar entry for the reservation in the user's calendar.

[0128] In some embodiments, the task flow processor 336 employs the assistance of a service processing module 338 ("service processor") to complete a task requested in the user input or to provide an informational answer requested in the user input. For example, the service processor 338 can act on behalf of the task flow processor 336 to make a phone call, set a calendar entry, invoke a map search, invoke or interact with other user applications installed on the user device, and invoke or interact with third party services (*e.g.* a restaurant reservation portal, a social networking website, a banking portal, etc.). In some embodiments, the protocols and application programming interfaces (API) required by each service can be specified by a respective service model among the services models 356. The service processor 338 accesses the appropriate service model for a service and generates requests for the service in accordance with the protocols and APIs required by the service according to the service model.

[0129] For example, if a restaurant has enabled an online reservation service, the restaurant can submit a service model specifying the necessary parameters for making a reservation and the APIs for communicating the values of the necessary parameter to the online reservation service. When requested by the task flow processor 336, the service processor 338 can establish a network connection with the online reservation service using the web address stored in the service model, and send the necessary parameters of the reservation (*e.g.,* time,

date, party size) to the online reservation interface in a format according to the API of the online reservation service.

[0130] In some embodiments, the natural language processor 332, dialogue processor 334, and task flow processor 336 are used collectively and iteratively to deduce and define the user's intent, obtain information to further clarify and refine the user intent, and finally generate a response (*i.e.,* an output to the user, or the completion of a task) to fulfill the user's intent.

[0131] In some embodiments, after all of the tasks needed to fulfill the user's request have been performed, the digital assistant 326 formulates a confirmation response, and sends the response back to the user through the I/O processing module 328. If the user request seeks an informational answer, the confirmation response presents the requested information to the user. In some embodiments, the digital assistant also requests the user to indicate whether the user is satisfied with the response produced by the digital assistant 326.

[0132] More details on the digital assistant can be found in the U.S. Utility Application No. 12/987,982, entitled "Intelligent Automated Assistant", filed January 18, 2010, U.S. Utility Application No. 61/493,201, entitled "Generating and Processing Data Items That Represent Tasks to Perform", filed June 3, 2011.

[0133] In most scenarios, when the digital assistant receives a user input from a user, the digital assistant attempts to provide an appropriate response to the user input with as little delay as possible. For example, suppose the user requests certain information (*e.g.,* current traffic information) by providing a speech input (*e.g.,* "How does the traffic look right now?"). Right after the digital assistant receives and processes the speech input, the digital assistant optionally provides a speech output (*e.g.,* "Looking up traffic information...") acknowledging receipt of the user request. After the digital assistant obtains the requested information in response to the user request, the digital assistant proceeds to provide the requested information to the user without further delay. For example, in response to the user's traffic information request, the digital assistant may provide a series of one or more discrete speech outputs separated by brief pauses (*e.g.,* "There are 2 accidents on the road. <Pause> One accident is on 101 north bound near Wipple Avenue. <Pause> And the second accident is on 85 north near 280."), immediately after the speech outputs are generated.

[0134] For the purpose of this specification, the initial acknowledgement of the user request and the series of one or more discrete speech outputs provided in response to the user request are all considered sub-responses of a complete response to the user request. In other words, the digital assistant initiates an information provision process for the user request upon receipt of the user request, and during the information provision process, the digital assistant prepares and provides each sub-response of the complete response to the user request without requiring further prompts from the user.

[0135] Sometimes, additional information or clarification (*e.g.,* route information) is required before the requested information can be obtained. In such scenarios, the digital assistant outputs a question (*e.g.*, "Where are you going?") to the user asking for the additional information or clarification. In some embodiments, the question provided by the digital assistant is considered a complete response to the user request because the digital assistant will not take further actions or provide any additional response to the user request until a new input is received from the user. In some embodiments, once the user provides the additional information or clarification, the digital assistant initiates a new information provision process for a "new" user request established based on the original user request and the additional user input.

[0136] In some embodiments, the digital assistant initiates a new information provision process upon receipt of each new user input, and each existing information provision process terminates either (1) when all of the sub-responses of a complete response to the user request have been provided to the user or (2) when the digital assistant provides a request for additional information or clarification to the user regarding a previous user request that started the existing information provision process.

[0137] In general, after a user request for information or performance of a task is received by the digital assistant, it is desirable that the digital assistant provides a response (*e.g.,* either an output containing the requested information, an acknowledgement of a requested task, or an output to request a clarification) as promptly as possible. Real-time responsiveness of the digital assistant is one of the key factors in evaluating performance of the digital assistant. In such cases, a response is prepared as quickly as possible, and a default delivery time for the response is a time immediately after the response is prepared.

[0138] Sometimes, however, after an initial sub-response provided immediately after receipt of the user input, the digital assistant provides the remaining one or more sub-responses one at a time over an extended period of time. In some embodiments, the information provision process for a user request is stretched out over an extended period of time that is longer than the sum of the time required to provide each sub-response individually. For example, in some embodiments, short pauses (*i.e.*, brief periods of silence) are inserted between an adjacent pair of sub-responses (*e.g.*, a pair of consecutive speech outputs) when they are delivered to the user through an audio-output channel.

[0139] In some embodiments, a sub-response is held in abeyance after it is prepared and is delivered only when a predetermined condition has been met. In some embodiments, the predetermined condition is met when a predetermined trigger time has been reached according to a system clock and/or when a predetermined trigger event has occurred. For example, if the user says to the

digital assistant "set me a timer for 5 minutes," the digital assistant initiates an information provision process upon receipt of the user request. During the information provision process, the digital assistant provides a first sub-response (*e.g.,* "OK, timer started.") right away, and does not provide a second and final sub-response (*e.g.,* "OK, five minutes are up") until 5 minutes later. In such cases, the default delivery time for the first sub-response is a time immediately after the first sub-response is prepared, and the default delivery time for the second, final sub-response is a time immediately after the occurrence of the trigger event (*e.g.,* the elapse of 5 minutes from the start of the timer). The information provision process is terminated when the digital assistant finishes providing the final sub-response to the user. In various embodiments, the second sub-response is prepared any time (*e.g.,* right after the first sub-response is prepared, or until shortly before the default delivery time for the second sub-response) before the default delivery time for the second sub-response.

[0140] In some embodiments, the digital assistant allows the user to submit additional user requests while the information provision process for a previously received user request is still underway. As a result, multiple concurrent information provision processes are maintained between the digital assistant and the user. In some embodiments, the same digital assistant serves multiple users present at the same location (*e.g.,* in the same room or the same vehicle) or sharing the same set of I/O devices (*e.g.,* speakers, displays, microphones, keyboards, etc.). As a result, multiple concurrent information provision processes are maintained between the digital assistant and the multiple users using the same set of shared I/O devices.

[0141] In general, when multiple concurrent information provision processes are maintained between the digital assistant and the user or multiple users sharing the same set of I/O devices, multiple responses may become concurrently available to be delivered to the user or multiple users. For example, sometimes, a user or multiple users sharing the same set of I/O devices may issue multiple requests to the digital assistant via several discrete speech inputs within a short period of time. Sometimes, digital assistants are capable of handling multiple user requests from the same user in parallel and preparing the appropriate response for each of the multiple user inputs in parallel. If the digital assistant receives one or more subsequent speech inputs before the digital assistant has completed an information provision process initiated by an earlier speech input received from the user, the digital assistant will accumulate multiple responses in the delivery pipeline. In some embodiments, a response or sub-response considered available for delivery when all the data needed to formulate the output object (*e.g.,* UI objects, speech output, sound, etc.) has been delivered to the I/O module from one or more internal components of the user device and/or from one or more remote servers. In some embodiments, a response or sub-response is considered available for delivery when its corresponding output objects (*e.g.,* UI objects, speech output, sound, etc.) have actually been completely formulated and ready to be presented to the user through one or more output devices (*e.g.,* a speaker, a display, etc.).

[0142] Sometimes, responses to multiple user requests received at different times become concurrently available because the responses took different amount of time to generate. Sometimes, the information provision process initiated by an earlier user request includes multiple discrete sub-responses to be provided to the user one at a time over an extended period of time. Therefore, before all the sub-responses are provided to the user, responses to one or more subsequently received user requests may become concurrently available to be provided to the user. The multiple concurrently available responses include a response or sub-response to a first user request (*e.g.,* the earlier received user request), and a response or sub-response to a second user request (*e.g.,* at least one of the subsequently received user requests).

[0143] In some embodiments, a subsequent user request is referred to as a "user-initiated interruption" to the information provision process initiated by an earlier received user request. In some embodiments, the subsequent user request is only considered as a "user-initiated interruption" to the information provision process initiated by the earlier received user request if concurrently available responses and sub-responses of the previous and subsequent user requests cannot or should not be delivered to the user simultaneously (*e.g.,* on the same display or over the same audio channel). In some embodiments, the subsequent user request is only considered as a "user-initiated interruption" to the information provision process initiated by the earlier received user request if concurrently available responses and sub-responses of the previous and subsequent user requests need to be delivered using the same audio output channel as speech outputs.

[0144] In some conventional systems, the digital assistant implements a default ordering rule to output responses prepared by concurrent information provision processes for two or more user requests. In some embodiments, the digital assistant implements a first-in-first-out (FIFO) approach for handling user-initiated interruptions at the request level. According to the FIFO approach on the request level, no response to a subsequently received user request is provided until all of the responses to all previously received user requests have been provided to the user. In some embodiments, the default rule adopts a last-in-first-out (LIFO) approach on the request level. When a LIFO approach on the request level is used, the digital assistant always respond to the last received user request first and either abandon the incomplete information provision processes for the previously user requests or return to them after a complete response has been provided for the last user request.

[0145] Other than the user-initiated interruptions described above, sometimes, interruptions may be initiated by programmed actions previously established by the user, the digital assistant, and/or third-party applications or systems executing on or communicating with the user device. These interruptions are referred to as "system-initiated interruptions." For example, sometimes, the user may demand or permit the digital assistant to generate a reminder or notification based on occurrence of certain trigger events. When the digital assistant detects that the predetermined trigger event(s) have occurred, the digital assistant or the user device generates and delivers an alert item (*e.g.,* a speech output, an alert sound, an alert message, a popup banner, badge, or message) providing content of the reminder or notification to the user. For example, an alert item for a previously established meeting reminder would be provided to the user when the reminder time specified in the meeting reminder has been reached. For another example, when an instant message or telephone call is received, a notification or alert for the instant message or telephone call is provided to the user without any delay.

[0146] In conventional systems, a default rule allows an alert item for a reminder or notification to be provided to the user as soon as its trigger event has occurred. In some conventional systems, the default rule is used regardless of whether the user is engaged in any other activity or if the digital assistant and the user are engaged in an existing conversation. In some conventional systems, the user is allowed to alter the default rule beforehand in a user preference setting to always prevent reminders or notifications (or a sub-category thereof) to be provided at all (*e.g.,* by turning off reminders or notification for particular applications or events) until the user preference setting is manually altered by the user again.

[0147] Although the default rules for handling delivery of outputs generated by concurrent information provision processes, reminders, and push-notifications are suitable in many scenarios, the default rules do not work well in all contexts. Even though the user can sometimes modify the default rules by changing a preference setting beforehand, the preference setting selected beforehand is merely a new default rule replacing an earlier default rule. These default rules still do not always work well since they do not respond intelligently under different contexts. Therefore, it would be advantageous to implement a case-by-case and context-sensitive way of handling interruptions initiated by the user, the digital assistant, the operating system, and/or third-party applications or processes not controlled by the digital assistant.

[0148] As will be described in more details later in this specification, a context-sensitive interruption handler (*e.g.,* the interruption handling module 340 in FIG. 3A) is implemented on top of the default rules for providing responses to the user requests and/or for providing the alert items for reminders and notifications. In some embodiments, the interruption handler gathers information regarding the present context in real-time, and determines in real-time whether the default rules for prioritizing deliveries of responses, reminders and/or notifications need to be altered such that a more appropriate ordering of the deliveries is used. For example, in some contexts, it would be more suitable to delay responding to a later received input, while other times it may be more suitable to delay providing the response to the earlier received input. In addition, in some contexts, it is more suitable to delay providing a reminder or notification even if the trigger event(s) for the reminder or notification have occurred, while in others, it is more suitable to provide the reminder or notification as soon as the trigger event(s) have occurred. In addition, sometimes, if it is best not to alter the default delivery time of at least one of the responses, while in others, it is more suitable to alter the delivery times for either or both responses in order to accommodate the delivery priorities determined based on the present context.

[0149] In some embodiments, the context-sensitive interruption handler of the digital assistant intercepts the responses, reminders, and/or notifications before they are provided to the user, and determines dynamically in real-time, a relative urgency between the responses, reminders, and/or notifications. The context-sensitive interruption handler of the digital assistant then provides the responses, reminders, and/or notifications in an order based on the relative urgency thereof. In some embodiments, since the context may change again during the time it takes for the most highly prioritized response/reminder/notification to be provided to the user, the relative urgency is re-evaluated among the remaining and any newly available responses, reminders, and notifications. In some embodiments, the re-evaluation takes into account new information that alters the present context.

[0150] In some embodiments, the interruption handler is invoked and the relative urgency evaluation is only performed for concurrently available responses, reminders, and notifications that are not suited to be provided concurrently through the same output channel (*e.g.,* the audio interface). For example, if a reminder can be provided via a graphical interface, and a response to user input can be provided to the user via a speech output, the digital assistant can optionally provide the reminder and the response simultaneously using the graphical interface and the speech output without resorting to the interruption handler.

[0151] In some embodiments, the digital assistant invokes the interruption handler only when the digital assistant is operated in a hands-free and/or eyes-free mode. In some embodiments, the digital assistant prioritizes the concurrently available outputs (*e.g.,* responses, reminders, and/or notifications) for delivery one at a time over a single output channel when the digital assistant detects that the user is likely to have diminished or impaired ability to focus on multiple output channels at the same time.

[0152] Figure 4 is a flow diagram of an exemplary process 400 implemented by a digital assistant to provide

concurrently available responses in an order that is based on the relative urgency between the concurrently available responses. In some embodiments, the relative urgency is determined dynamically by an interruption handler (*e.g.,* the interruption handler 340 in FIG. 3A) of the digital assistant in real-time based on the present-context.

[0153] In the exemplary process 400, a first input is received (402) from the user. In some embodiments, the first input is a first speech input received from the user. For example, a user speaks to the digital assistant to provide a first user request. In some embodiments, the first input is a user request submitted through one or more of multiple input channels, such as a selection of a user interface object (*e.g.,* a search button on a touch-sensitive display), a textual command entered via a keyboard, invocation of a mechanical controller (*e.g.,* a doorbell of a house coupled to a digital assistant, or a mouse button), and so on.

[0154] Upon the first input being received from the user, the digital assistant initiates (404) a first information provision process in response to receipt of the first input. In some embodiments, the first information provision process includes generating at least a first response and a second response to the first input. In some embodiments, the first response and the second response are a first sub-response and a second sub-response of a complete response to the same user input. Initiating the information provision process does not necessarily mean that the process is completed, *i.e.,* the generation of the first and second responses may not both be completed, and the first response and the second responses may not both be provided to the user. In some embodiments, the first response and the second response are two discrete navigation instructions (*e.g.,* turn-by-turn directions) generated in response to a single navigation request received from the user. In some embodiments, the two discrete navigation directions are to be provided to the user as two discrete speech outputs at two different waypoints of a route during navigation (*e.g.,* driving or walking). In some embodiments, the first response and the second response are two consecutive sub-responses of a series of discrete sub-responses to the first input. The first information provision process includes provision or delivery of the series of discrete sub-responses to the user one at a time over an extended period of time without requiring further prompts from the user. In some embodiments, provision of all of the series of discrete sub-responses terminates the first information provision process.

[0155] In some embodiments, the first response and the second response are two discrete speech outputs reading two distinct information items in a list of information items retrieved by the digital assistant in response to the same user request. For example, in some embodiments, if the user request is for reading a cooking recipe, the first response and the second response are two discrete speech outputs reading two different (*e.g.,* consecutive) steps of the cooking recipe. In some embodiments, if the user request is for reading a list of newly received email messages, the first response and the second response are two discrete speech outputs reading two different email messages identified by the digital assistant in response to the user request. In some embodiments, if the user request is for reading a particular email message, the first response and the second response are two speech outputs reading two sub-parts of the single email message, *e.g.,* a header and a message body of the single email message identified by the digital assistant in response to the user request.

[0156] In some embodiments, the first input is a search request, and the first response and the second response are speech outputs reading two different search results retrieved in response to the search request. In some embodiments, the first input is a list-reading request, and the first response and the second response are speech outputs reading content of two different information items in a list of information items identified by the user. In some embodiments, the list-reading request is a request to read a cooking recipe, a list of email messages, a list of search results, a list of instructions, a list of diagnostic procedures, a list of exercise routines, a list of calendar entries, a list of reminders, a list of turn-by-turn directions, a list of SMS messages, a list of voice mail messages, a list of passages from a book or article, or the like. In some embodiments, the first response and the second responses are discrete speech outputs reading different information items (*e.g.,* different ingredients in a recipe, different email messages, different search results, different steps in the instructions or procedures, routines, different calendar or reminder entries, different passages, etc.) from the list. In some embodiments, depending on the nature of the information items in the list, the first response and the second response are either delivered consecutively with a short pause in between, or delivered upon occurrence or respective trigger events (*e.g.,* arrival of particular trigger times, or arrival at particular trigger locations, etc.)

[0157] In some embodiments, the first response is a leading summary of one or more additional sub-responses including the second response, and the second response provides the details of an information item requested by the first user input. For example, in some embodiments, if the user request is searching for restaurants nearby, the first response is a speech output summarizing all the retrieved search results (*e.g.*, "I found 5 restaurants nearby."), and the second response is a speech output detailing a particular search result (*e.g.,* "The nearest one is five miles south of here."). In some embodiments, the first response is a speech output conveying an acknowledgement or rephrase of the user request (*e.g.*, "OK, searching for nearby restaurants...") and the second response is a speech output (*e.g.,* "No restaurant is found within 100 miles.") containing the information retrieved by the digital assistant in response to the user request.

[0158] In some embodiments, the first response and the second response are logically discrete portions of a complete response to a single user request that are to be provided serially in time without further prompts from the user. In some embodiments, in a default scenario where no interruption (*e.g.,* an intervening user input, reminder, or notification) is detected, provision of the first and the second responses proceeds serially in time with a minimum amount of delay in between. In some embodiments, in a default scenario where no interruption (*e.g.,* an intervening user input, reminder, or notification) is detected, the first response and the second response are provided serially with a short pause in between. In some embodiments, the duration of the pause is comparable to a pause used in normal human speech for separating discrete semantic concepts embodied in consecutive clauses or phrases. In some embodiments, in a default scenario where no interruption (*e.g.,* an intervening user input, response to an intervening input, reminder, or notification) is detected, the first response and the second response are provided upon occurrence of respective predetermined triggering events (*e.g.,* close proximity to respective waypoints on a route, arrival of predetermined times, and so on). In some embodiments, the digital assistant dynamically determines the appropriate times to deliver the first response and the second response to the user by monitoring a number of sensors (*e.g.,* a location sensor such as a GPS, a speech sensor, a compass, a system clock, etc.) coupled to the user device.

[0159] After the first information provision process is initiated in response to receipt of the first input from the user, the digital assistant provides (406) the first response to the user. In some embodiments, the first response is a speech output among a series of speech outputs to be provided to the user in response to the user request. In some embodiments, the first response is any one of the series of speech outputs other than the last one of the series of speech outputs. In some embodiments, the first response and the second response are any two responses that are either consecutive responses in a series of responses or separated by one or more other responses.

[0160] In some embodiments, the first input is a request to establish a reminder to be triggered at a later time by occurrence of a specified trigger event (*e.g.,* arrival of a specified reminder time that is one hour before a specified event time). In such embodiments, the first response is an acknowledgement to the first input or a confirmation that the reminder has been set. The second response is an alert item to be delivered to the user at the later time when the specified trigger event occurs. In some embodiments, the digital assistant does not consider the establishment of a reminder and the provision of the alert item a continuous information provision process (e.g., the first information provision process), and the interruption handling described in this specification does not apply. This exclusion is advantageous in some embodiments because it eliminates the digital assistant's need to maintain the information provision process for a long period of time.

[0161] In some embodiments, the user can interrupt the digital assistant while the digital assistant is in the process of providing the series of responses for first user input. In some embodiments, after or concurrent with the provision of the first response to the user, but before provision of the second response to the user, the digital assistant detects (408) an event operable to initiate a second information provision process. In some embodiments, the event that is operable to initiate a second information provision process is a second input received from the user. In this scenario, the second input is a user-initiated interruption to the first information provision process. In some embodiments, the second input received from the user is a second speech input.

[0162] In various embodiments, the digital assistant can capture a user-initiated interruption in a number of different ways. For example, in some embodiments, the digital assistant maintains a continuous listening mode. Therefore, even while the digital assistant is in the middle of delivering one of the responses (*e.g.,* the first response) to the first input to the user, the digital assistant is able to capture new speech inputs uttered by the user. For example, the user may provide a second speech input (*e.g.,* "Remind me to pick up my dry cleaning today") when the digital assistant is in the middle of providing the first response ("The nearest one is Tammy Cafe, 0.2 miles, cash only, serves ...") to the first user input (*e.g.,* a first speech input "Find restaurants nearby.").

[0163] In some embodiments, the digital assistant enters into the listening mode only during the brief pause between consecutive speech outputs provided to the user. In some embodiments, the digital assistant enters into and remains in the listening mode when the digital assistant is silent between consecutive responses provided to the user. In some embodiments, the silence between responses occurs without special arrangement because the trigger events for each of the responses naturally take a certain amount of time to occur. While the digital assistant is in the listening mode, if the user provides a second speech input, the digital assistant captures the second speech input after or concurrent with the provision of the first response to the user, but before provision of the second response to the user. For the purposes of the interruption handling process, it is presumed or required that the receipt of the second input does not alter validity of the responses (e.g., the second response) to the first input that have not been provided to the user yet.

[0164] In some embodiments, the event that is operable to initiate a second information process is a reminder or notification generated based on detection of some predetermined trigger events by the user device, the digital assistant, or a third-party application or system. For example, if the user has previously set a reminder or alarm clock to go off at a predetermined reminder time (*e.g.,* a reminder to call mom at 2pm today), an event that is operable to initiate a second information provision proc-

ess is the arrival of the 2pm as indicated by a system clock. In some embodiments, the information provision process initiated by the event includes preparation and provision of a speech and/ or visual alert item delivered to the user about the content of the reminder. For example, a speech output saying "OK, it is time to call your mom" is a response to the event (*e.g.,* the arrival of the reminder time). If the user has specified to have the reminder go off 5 minutes before the scheduled call time (*i.e.,* the specified event time), the trigger event is the arrival of 5 minutes before the scheduled call time. For example, in some embodiments, the digital assistant provides a speech output that says, "It's five minutes to 2pm now. You have scheduled a call with your mom at 2pm."

[0165] In some embodiments, other types of events are operable to initiate a second information provision process. For example, in some embodiments, a notification process (*e.g.,* "push notification") initiated by a third-party application or system is permitted on the user's device. If the user accepts push notifications for arrival of emails, instant messages, or product updates, etc., an event that is operable to initiate a second information provision process is the arrival of such a push notification. The second information provision process initiated by such an event includes preparation and delivery of an alert item containing the content of the notification to the user. In some embodiments, the alert item is provided to the digital assistant by the third-party application or system to forward to the user. In some embodiments, the event does not include generation and delivery of an alert item by a third-party application or process not currently controlled by the digital assistant. In other words, the notification generated by third party applications not controlled by the digital assistant can supersede the digital assistant's interruption handling and deliver the notification without regard the digital assistant. In some embodiments, the digital assistant intercepts the push notification before it is delivered to the user through applications or the operating system not currently controlled by the digital assistant. In some embodiments, the digital assistant only detects event that are likely to use the same output channel as the first information provision process.

[0166] In some embodiments, in response to the detection of the event that is operable to initiate the second information provision process, the digital assistant initiates (410) the second information provision process. In some embodiments, the second information provision process includes preparing and delivering a third response to the event. In some embodiments, if the event is the receipt of a second speech input requesting information, preparing a third response to the detection of the event includes generating a respective speech output to provide the information requested by the second speech input. In some embodiments, the third response is a complete response to the second speech input. In some embodiments, the third response is one of a series of sub-responses to the second speech input to be provided to the user over an extended period of time.

[0167] In some embodiments, the event is the occurrence of a predetermined trigger event for a previously established reminder, the second information provision process includes generating an alert item (*e.g.,* a speech output) providing the content of the reminder to the user. In some embodiments, the third response is a speech output providing content of the reminder.

[0168] In some embodiments, the event is the arrival of a push-notification, and the third response is a speech output providing content of the push notification. For example, in response to receipt of an email, the digital assistant can generate a speech output that says "New email from Tae-Woong about 'Funny photos. "' For another example, upon receipt of an SMS message, the digital assistant generates a speech output that says "New SMS message from Peter saying "Those pictures lie!".

[0169] In some embodiments, the digital assistant detects that there are concurrently available responses triggered by two or more different sources waiting to be provided to the user. The digital assistant initiates an interruption handling process to determine how to intelligently handle the interruptions rather than resorting to the default rules that deterministically decide which response to provide to the user first without consideration of the present context. In some embodiments, the present context is determined based on a combination of multiple factors. In some embodiments, the particular set of factors relevant to form the present context is not necessarily static, and the present values of the factors are also not necessarily static. In some embodiments, for different type of events and responses, a different set of priority parameters are considered in evaluating the relative priority or urgency for delivery. Therefore, the outcome of the context determination by the interruption handling process often varies from case to case.

[0170] In some embodiments, after the first information provision process and the second information provision processes are both initiated and uncompleted (*i.e.,* a complete response has not been provided according to either the first or the second information provision process), the digital assistant determines (412) a relative urgency between the second response and the third response. In some embodiments, the digital assistant determines the relative urgency based on the present context. In some embodiments, the digital assistant gathers the context information for the present time after detecting the concurrent availability of both the second and the third responses. In some embodiments, the digital assistant consults a database of rules and priority parameters to assess the relative urgencies of the second and the third responses. In some embodiments, the digital assistant selects one of multiple different sets of priority factors for evaluating the relative urgency between two responses based on the content of the two responses and/or the domain associated with the two responses. In some embodiments, the digital assistant selects one or more additional priority factors for consideration when

outcome for relative urgency determination based on an initial set of priority factors is not conclusive.

[0171] In some embodiments, the digital assistant determines whether it is currently operating in a hands-free mode or an eyes-free mode when concurrent responses for multiple information provision processes are available. In some embodiments, the digital assistant only invokes the relative urgency determination when the digital assistant is operating in the hands-free or eyes-free mode. In some embodiments, the digital assistant detects that the user is in motion, and invokes the hands-free or eyes-free mode upon detecting that the user is motion. In some embodiments, the digital assistant detects that the user is currently performing an action that is likely to cause impaired attention to the visual interfaces of the user device, such as navigating a vehicle, walking, jogging, exercising, operating a device or machinery, operating an application or device not currently controlled by the digital assistant (*e.g.,* a gaming device or game not controlled by the digital assistant).

[0172] In some embodiments, the digital assistant determines whether the second response and the third response are suitable to be provided in parallel on different output channels of the user device. In some embodiments, the digital assistant invokes the relative urgency determination upon determination that the second response and the third response are not suited to be provided in parallel on different output channels of the use device. In some embodiments, the digital assistant does not invoke the relative urgency determination unless it has determined that the second response and the third response are not suited to be provided in parallel on different output channels of the use device.

[0173] In some embodiments, the digital assistant determines whether it is operating in a hands-free or eyes-free mode. In some embodiments, the digital assist determines that the second response and the third response are not suited to be provided in parallel on different output channels if it determines that the digital assistant is currently operating in the hands-free or eyes-free mode.

[0174] In some embodiments, after the relative urgency is determined, the digital assistant proceeds to provide (414) one of the second response and the third response in an order based on the determined relative urgency. In some embodiments, after one of the second response and the third response is provided to the user, the digital assistant resumes providing the remaining response. In some embodiments, the digital assistant detects that there are additional pending sub-responses for the first and the second inputs to be provided to the user. In some embodiments the digital assistant performs the relative urgency evaluation between the remaining concurrently available responses for the two inputs based on the new context. For example, if the user has been traveling along a route, the present location and time are constantly changing, and the content associated with the next pair of concurrently available responses are also changing. Therefore, the relative urgency is determined dynamically again. Then, one of the remaining responses for both the second input and the first input is provided based on the determined relative urgency.

[0175] For example, between the second response and the third response, if the second response is determined to have a higher relative urgency, the second response is provided first. If there are additional responses for the first input, *e.g.,* a fourth response, then the fourth response is compared against the third response based on the newly updated context. If the third response is determined to have a higher relative urgency than the fourth response, then the third response is provided before the fourth response. The process may continue until both the first information provision process and the second information provision process are completed.

[0176] As set forth earlier, the interruption handling process alters the default rules for ordering and timing response deliveries to the user. In some embodiments, after the a series of discrete sub-responses to the first user input have been prepared, the digital assistant determines a respective default time for delivering each of the series of sub-responses to the first user input. In some embodiments, the digital assistant dynamically overrides at least one of the respective default delivery times for delivering the second and the third responses. For example, the digital assistant may decide to move the actual delivery time(s) for either or both the second and third responses to either before or after their respective default delivery time(s) depending on the context and the determined relative urgency.

[0177] For example, suppose the second response is a navigation instruction to be delivered 30 seconds in the future, and the third response is a response to a weather information request to be delivered 10 seconds in the future. Suppose that it takes the user 30 seconds to safely maneuver the turn specified by the second response, and the third response lasts 40 seconds. Based on these data, the digital assistant determines that the navigation instruction is more urgent and should be delivered first, so that the user does not miss the turn. Furthermore, the digital assistant also decides to move the delivery time of the weather report to 60 seconds from now, such that the user can safely maneuver through the turn and not be distracted by the weather information response. In some embodiments, the digital assistant generates a time filler (*e.g.,* a speech output "hold on, turn coming up ...") if the response having the higher urgency is not going to be delivered immediately. In some embodiments, the time filler is a speech output (*e.g.,* "hold on, turn coming up...") generated based on the response having the higher urgency. In some embodiments, the time filler is a generic speech output (*e.g.,* "hold on ...")

[0178] In some embodiments, the first input is a directions request, and the second response is a particular navigation instruction having a default delivery time associated with the current location of the user relative to the waypoint associated with the navigation. In some em-

bodiments, the digital assistant determines the respective default delivery time for the navigation instruction based on a predetermined proximity (*e.g.,* 1 mile, 50 yards, or 30 feet) between the current location of the user and the respective waypoint associated with the navigation instruction along a predetermined route.

[0179] As set forth earlier, in some embodiments, the digital assistant determines a present context associated with the user and determines the relative urgency between the second response and the third response based on the present context associated with the user. In some embodiments, the present context is formed by a combination of one or more factors or priority parameters including: the current location of the user, the current speed of the user, the current travel direction of the user, a location (*e.g.,* waypoint or trigger location) associated with the second response, and a location (*e.g.,* waypoint or trigger location) associated with the third response. In some embodiments, the digital assistant determines whether the second response is associated with a location within a predetermined distance from the current location of the user. In some embodiments, the predetermined distance is determined based on the current speed of the user. In some embodiments, the digital assistant determines the relative urgency between the second response and the third response based on whether the user is likely to pass by a location associated with the second response within a predetermined time window from the current time. In some embodiments, the predetermined time window is based on sizes of respective time periods (*e.g.,* time buffers) required to deliver the second response and the third response. In some embodiments, upon determining that the user is likely to pass by the location associated with the second response within the predetermined time window from the current time, the digital assistant assigns a higher level of urgency to the second response than the third response. In some embodiments, upon determining that the user is unlikely to pass by the location associated with the second response within the predetermined time window from the current time, the digital assistant assigns a lower level of urgency to the second response than the third response. In some embodiments, the digital assistant monitors the current location, the current direction/heading, and the current speed of the user, *e.g.,* using a GPS device, and determines the present context based on the monitoring.

[0180] In some embodiments, the time period (*e.g.,* a time buffer) required to deliver a response to the user is very short, such as a short alert sound with no accompanying speech outputs. In some embodiments, if the response is a speech output, the time period required to deliver the response is the duration of the speech input plus an arbitrary short pause (*e.g.* 50ms or Is) before and/or after the speech output. In some embodiments, the time period required to deliver the response also includes the time for the user to carry out the action specified in the response. For example, if the response is a speech output providing a turn-by-turn direction or a instruction regarding a yoga pose, the time period (*e.g.,* a time buffer) required to deliver the response includes the duration of the speech output plus a period of time for the user the make the turn or form the yoga pose. In some embodiments, the digital assistant uses the time period(s) required to deliver the response(s) to determine whether and how to alter (*e.g.,* advance or delay) the default delivery time of the response(s).

[0181] As described earlier, in some embodiments, the second response is provision of an information item in response to an information request, while the third response is an alert item generated for a previously established reminder or arrival of a push notification. In some embodiments, the digital assistant determines whether delivery of the second response at its respective default delivery time is likely to adversely affect a utility of the reminder or notification to the user. For example, suppose the reminder is for calling a customer at 2pm, and the reminder has a default reminder time of 5 minutes to 2pm. Further suppose that at 2:54pm, the digital assistant started an email reading session in response to a user request (*e.g.,* the first input), and a first response ("OK, you got 5 new emails.") acknowledging the user request has been provided immediately. The second response scheduled to be delivered next is a speech output reading the first email. The default delivery time for the second response is 2 seconds after the delivery of the first response, but the second response requires a time buffer of 1 minute to finish. Therefore, the alert item would be overdue before the second response can be completely delivered. In this scenario, the digital assistant determines whether finishing delivering the second response and pushing the delivery time of the alert item by a few seconds is likely to affect the utility or validity of the alert item. In this example, the digital assistant determines that pushing the delivery time of the alert item by a few seconds would not adversely affect the utility or validity of the alert item. Consequently, the digital assistant assigns a higher relative urgency to the second response, and delays the delivery of the alert item until after the second response has been delivered to the user.

[0182] Continuing with the above example, suppose that after the delivery of the second response, the digital assistant is faced with a fourth response (*e.g.,* a speech output reading the next one of the 5 emails) with a default delivery time 2 seconds after the delivery of the second response. The digital assistant then determines whether finishing delivering the fourth response and delaying the delivery time of the alert item by the time buffer of the fourth response (*e.g.,* 1 minute) is likely to affect the utility or validity of the alert item. In this example, the digital assistant determines that delaying the delivery time of the alert item by another 1 minute would adversely affect the utility or validity of the alert item. Consequently, the digital assistant assigns a higher relative urgency to the alert item, and delays the delivery of the fourth response until after the alert item has been delivered to the user.

After the delivery of the alert item, the user can proactively terminate the email reading session with another input so as to get ready for the phone call.

[0183] As described above in various examples, the digital assistant detects whether there is a timing conflict between deliveries of the second response and the third response based on respective default delivery times of the second response and the third response. In some embodiments, the digital assistant determines the relative urgency between the second response and the third response upon detection of the timing conflict. In some embodiments, the digital assistant does not initiate the determination of the relative urgency between the second response and the third response unless it has detected the timing conflict. In some embodiments, the digital assistant overrides the respective default delivery time for at least one of the second and third responses based on the determined relative urgency. In some embodiments, the digital assistant determines whether the default delivery time of a particular can be moved without impacting the utility or validity of the response. If the default delivery time of one particular response (*e.g.,* turn by turn direction, or real-time cooking instruction, or real-time chemistry experiment instruction etc.) cannot be moved without impacting its utility or validity of the instructions, the digital assistant seeks to move (*e.g.,* advance or delay) the default delivery time of the other response. In some embodiments, the default delivery time of a response is a projected delivery time for the response as if the response were the only response waiting to be delivered to the user. In some embodiments, the digital assistant determines the relative flexibility in the default delivery times of the second response and the third response based on the present context. In some embodiments, the digital assistant determines the relative delivery priority and/or adjusted delivery times of the second response and the third response based on the determined relative urgency and the relative flexibility in the default delivery times of the responses.

[0184] In some embodiments, other factors are considered when determining the present context associated with the user. For example, in some embodiments, an event type is determined for the event that initiated the second information provision process. For example, the event type includes a type based on whether the event is a speech input, a reminder, or a notification. In some embodiments, the speech event type further includes sub-types such as speech input that requires list reading or a single response. In some embodiments, the reminder event type further includes sub-types such as critical reminders, reminders requiring travel or preparation, reminders for to-dos. In some embodiments, the notification event type further includes sub-types divided by sources such as email notifications, SMS notifications, product updates notifications, emergency alert notifications, security breach notifications, routine broadcast notifications, etc. In some embodiments, the digital assistant uses the event type to determine the relative urgency between the second response and the third response. In some embodiments, the first input is also considered an event and is associated with various event types.

[0185] In some embodiments, the digital assistant also reviews the content of the second response and the third response to determine the relative urgency. For example, some keywords (*e.g.,* injury, emergency, 911, hurry, etc.) would indicate higher urgency, importance, and/or time criticalness. In some embodiments, the digital assistant scans the content of the second response and the third response to determine whether such keywords are present, and use the detected keywords as part of the present context.

[0186] In some embodiments, the digital assistant determines the amount of time that the second response and the third response will remain relevant given the current context. For example, a response about weather information is likely to remain relevant for hours, while a response about directions is unlikely to remain relevant for a long time when the user is driving. In addition, a response about weather information is likely to remain relevant for a longer period of time if the user is asking about the weather for the next day than if the user is asking about the weather for the current day right before he or she leaves for work. In some embodiments, the digital assistant takes into consideration of many different factors in determining the time period that the response is likely to remain valid and useful to the user.

[0187] In some embodiments, when determining the relative urgency of the second and the third responses, the digital assistant also considers the likely consequence of providing the second response later than the third response and/or the likely consequence of providing the third response later than the second response. For example, if the third response is an emergency advisory generated by the national security alert system and pushed to all networked devices in the country, the likely consequence of delaying that third response until after delivery of navigation instruction (*i.e.,* the second response in this example) is probably severe. On the other hand, if the third response is a notification about a product update, the likely consequence of delaying that notification until after delivery of the navigation instruction (*i.e.,* the second response in this example) is probably not severe. In the second case, delaying the navigation instruction would have more severe consequences since the user would likely miss a product update notification that is not time-sensitive. In some embodiments, the digital assistant uses the relative severity for delaying the responses to determine the relative flexibility in the default delivery times of the responses.

[0188] In some embodiments, the digital assistant allows the user to enter certain user preferences regarding the relative urgencies between responses. For example, in some embodiments, the digital assistant allows the user to enter a preference setting for setting navigation instructions to have a higher urgency if the user is traveling on an unfamiliar route and a lower urgency rat-

ing if the user has been on the route multiple times before. In some embodiments, whether the user is familiar with a region or route is determined based on a threshold number of times that the user has traveled on the route. In some embodiments, the preference setting is a qualitative setting provided by the user, and the digital assistant does not explicitly or quantitatively control the exact criteria for establishing when the preference is satisfied. In some embodiments, the digital assistant adapts the exact criteria for establishing when the preference is satisfied based on artificial intelligence techniques or crowd-sourced heuristics.

[0189] The above are merely some of the factors and priority parameters that the digital assistant can consider when forming the present context and evaluating the relative urgency of the concurrently available responses from multiple information provision processes. More details and examples regarding how the present context is used to determine the relative urgency between the second response and the third response are provided later in the specification.

[0190] As described earlier, the digital assistant sometimes enters a hands-free or eye-free mode of operation when it detects that the user is likely to have diminished abilities to focus on the display screen or use hand-operated user input devices, such as when the user is driving or operating another device. In some embodiments, when operating in hands-free mode or eyes-free mode, the digital assistant disables one or more input and/or output modes based on certain predetermined criteria (*e.g.*, current speech of the user). In some embodiments, when operating in hands-free mode, the digital assistant disables input channels that require movements of the user's hands. In some embodiments, when operating in hands-free mode, the digital assistant only receives user's request through one or more speech inputs provided by the user. In some embodiments, the digital assistant provides only audio outputs (*e.g.*, speech outputs) to the user when operating in eyes-free mode. In some embodiments, when operating in hands-free mode, the digital assistant provides speech outputs supplemented by visual information (*e.g.*, graphical or textual information) concurrently provided on a display screen. In some embodiments, the visual information persists on the screen after the speech input(s) have been provided to the user. In some embodiments, the digital assistant implements an eyes-free mode in which the digital assistant disables visual outputs (*e.g.*, textual and graphical outputs that are displayed to the user on a display screen) and provides all the pertinent information to the user via speech outputs. In some embodiments, when operating in eyes-free mode, the digital assistant also provides other audio outputs such as an audio alert, vibration alerts, and/or haptic feedbacks. In some embodiments, the digital assistant does not prohibit contemporaneous visual information being displayed along with audible outputs when operating in eyes-free mode. However, the audible outputs alone are sufficient to provide all the information the digital assistant needs to convey to the user.

[0191] As described above, in some embodiments, the digital assistant automatically initiates a hands-free mode and/or an eyes-free mode when the digital assistant detects that the user is navigating a vehicle. In some embodiments, the digital assistant detects that the user is navigating a vehicle based on a combination of several factors, *e.g.*, the current speed of the user, a sensor placed in a car seat that detects presence of a driver, a sensor in the vehicle ignition mechanism that detects ignition of the vehicle's engines, and/or detection of a mobile device connecting to the vehicle (*e.g.*, via Bluetooth or other wired or wireless connections). In some embodiments, the digital assistant initiates a hands-free mode and/or an eyes-free mode when the digital assistant detects that the user is in self-propelled motion such as swimming, jogging, running, walking, bicycling, rowing, and so on, based on preestablished motion patterns.

[0192] In some embodiments, the digital assistant initiates the hands-free mode and/or eyes-free mode when the user explicitly or implicitly requests the digital assistant to start the hands-free mode and/or the eyes-free mode. For example, the digital assistant initiates a hands-free mode and/or eyes-free mode when the user turns-off the display, turns on a power-saving mode, and/or enables a corresponding option (to enable hands-free mode or eyes-free mode) using any one of the multiple input modes (*e.g.*, speech, touch, and keyboard input modes).

[0193] As described above, in some embodiments, the digital assistant enables the hands-free mode and/or the eyes-free mode when the digital assistant detects that the user is likely engaged in one of a number of activities that occupy the user's hands and/or eyes. In some embodiments, the digital assistant detects whether the user is likely engaged in an activity that occupies the user's hands and/or eyes based on the application and/or application mode that the user is using on the user device. In some embodiments, when the user asks the digital assistant to initiate a training routine of a personal trainer application, the digital assistant determines that the user is likely to be engaged in an exercise routine and initiates the hands-free mode and/or eyes-free mode accordingly. In some embodiments, when the user asks the digital assistant to initiate a real-time walkthrough of a recipe, the digital assistant determines that the user is likely to be engaged in cooking according to the recipe, and initiates the hand-free mode and/or the eyes-free mode accordingly. In some embodiments, when the user asks the digital assistant to initiate a real-time walkthrough of a diagnostic procedure of another device (*e.g.*, a diagnostic and repair procedure of a home appliance), the digital assistant determines that the user is likely to be engaged in activities required by the diagnostic procedure, and initiates the hand-free mode and/or the eyes-free mode accordingly. Note, however, if the digital assistant determines that the diagnostic procedure is to be carried out on the user device, then digital assistant does not disable

the other input and output modes of the user device.

[0194] In some embodiments, the digital assistant enables the hands-free mode and/or the eyes-free mode when the digital assistant discovers that the display screen of the user's device is completely occupied by another application not controlled by the digital assistant. For example, if the digital assistant detects that the user is playing a video game or watching a movie in a full-screen mode using the device on which the digital assistant resides, the digital assistant initiates the hands-free and/or eyes-free mode to communicate with the user through speech and audio only.

[0195] In some embodiments, the digital assistant provides service to multiple users (*e.g.*, a primary user and one or more secondary users) at the same location (*e.g.*, sitting in the same vehicle or in the same room). In some embodiments, the digital assistant enables hands-free and/or eyes-free mode for communicating with one of the multiple users, if the digital assistant detects that a shared display screen of the user device is occupied by another one of the multiple users. For example, if a digital assistant is activated in a vehicle, and a built-in backseat display screen of the vehicle is occupied by a first passenger for playing a game or surfing the Internet, the digital assistant will use both the display screen and the speech-based interface to communicate with the first passenger while enabling the hands-free and/or eyes-free mode to communicate with the other passengers in the vehicle.

[0196] Exemplary process 400 in Figure 4 is a generally applicable process illustrating the operation of the interruption handling module. In some embodiments, the process 400 can be tailored for one or more particular domains (*e.g.*, vehicle navigation domain, live broadcast domain, cooking domain, physical therapy domain, exercise domain, academic instruction domain, medical diagnostic domain, technical support domain, etc.). In each of these domains, during an information provision process that lasts a long time and contains a series of multiple discrete sub-responses to a single user request, the digital assistant allows user-initiated interruptions and/or system-initiated interruptions. The digital assistant can intelligently and dynamically prioritize the delivery of responses from different concurrent information provision processes in real-time and on a case-by-case basis.

[0197] For example, FIG. 5A illustrate an example scenario in which the first information provision process is a process for providing navigation instructions (*e.g.*, turn-by-turn directions) to a user while the user is navigating a vehicle along a predetermined route to a predetermined destination. In this example, the digital assistant communicates with the user through speech inputs and outputs only.

[0198] As shown in FIG. 5A, at location O, the user requested directions to a library using a first speech input SI_1 ("Take me to the library."). The digital assistant receives the first speech input SI_1, and retrieves a series of navigation instructions (*e.g.*, turn-by-turn directions) to be delivered to the user at five different waypoints X_1-X_5 along a route. The digital assistant provides the first speech output SO_1 ("OK. Please proceed forward.") in response to the navigation request. Right before (*e.g.*, 30 feet before) the user reaches the first waypoint X_1 where the second speech input SO_2 ("Turn right at the next intersection.") would be provided, the user provides a second speech input SI_2 ("Is it about to rain?") to the digital assistant at time t_2 and location l_2. In response to the second speech input SI_2, the digital assistant prepares a third speech output SO_3 ("No, it is not going to rain today in Cupertino."). The digital assistant now has two concurrently available responses SO_2 and SO_3 from two different information provision processes waiting to be delivered to the user. The digital assistant evaluates the relative urgency between the second speech output SO_2 and the third speech output SO_3. To determine the relative urgency, the digital assistant assesses the current location of the user, the current speed of the user, the location of the waypoint associated with the second speech output SO_2, and the likely consequence of delivering the second speech output SO_2 before the third speech output SO_3, and vice versa. The digital assistant also considers the duration by which SO_2 and SO_3 are likely to remain valid. Since the user is already very close (*e.g.*, less than 30 feet away) to the waypoint X_1 at this point, SO_3 is likely to become invalid very soon, while SO_2 will remain valid for a long time still. The digital assistant further determines that not delivering SO_2 as soon as possible would cause the user to miss the turn, while not delivering SO_3 as soon as possible would have negligible ill consequences. As a result, the digital assistant determines that SO_2 has a higher relative urgency than SO_3, and proceeds to provide SO_2 right before the waypoint X_1. After delivery of SO_2, and waiting till the user has made the turn at X_1, the digital assistant proceeds to provide SO_3. At this point, the second information provision process terminates, and the first information provision process continues on.

[0199] Continuing with this example. Suppose the user continues on the route, and right before reaching the second waypoint X_2, the digital assistant provides a fourth speech output SO_4 ("Turn left at the next intersection.") After the user makes the left turn, and way before the user reaches the next waypoint X_3 associated with fifth speech output SO_5 ("Turn right at the next intersection."), the user provides a third speech input SI_3 ("Set a reminder to call mom at 2pm") which starts a third information provision process. In response to the third speech input SI_3, the digital assistant prepares a sixth speech output SO_6 ("OK, reminder set for 5min to 2pm"). At this time, the digital assistant again has two concurrently available responses SO_5 and SO_6 from two different information provision processes waiting to be delivered to the user. The digital assistant evaluates the relative urgency between the SO_5 and SO_6. To determine the relative urgency, the digital assistant assesses the current location, the current time, and the location and time associated

with SO_5 and SO_6. The digital assistant determines that neither SO_5 nor SO_6 are very urgent and that there is no timing conflict between SO_5 and SO_6. As a result, the digital assistant decides to follow the default delivery times for SO_5 and SO_6, and proceeds to deliver SO_6 ("OK, reminder set for 5min to 2pm") right way, and wait to deliver SO_5 ("Turn right here.") until the user has reached 30 feet of the waypoint X_3.

[0200] Continuing with the above example, the user then approaches the next waypoint X_4 associated with speech output SO_7 ("Turn right at the next intersection.") right before 5 minutes before 2pm. An alert item (*e.g.*, an eight speech output SO_8 "It's 5 to 2. Time to call mom.") has been generated for the reminder established earlier, and the default delivery time for the alert item is right now (i.e., 5 minute to 2pm). At this time, the digital assistant again has two concurrently available responses SO_7 and SO_8 from two different information provision processes waiting to be delivered to the user. The digital assistant evaluates the relative urgency between the SO_7 and SO_8. To determine the relative urgency, the digital assistant assesses the current location, the current time, and the location and/or time associated with SO_7 and SO_8. The digital assistant determines that SO_7 needs to be provided right way because the user is already very close (*e.g.*, less than 30 feet) to the waypoint X_4. The digital assistant also considers the time between the reminder time and the actual event time (2pm). The digital assistant optionally determines that there would be no preparation needed to call mom, and concludes SO_8 can be delayed until after delivery of SO_7 without adversely impacting the utility of SO_8 to the user. As a result, the digital assistant delivers SO_7 right way. After SO_7 has been delivered and the user has safely made the right turn, the digital assistant delivers SO_8 1 minute after its default delivery time, and 4 minutes before the specified event time. No harm is done by the delayed delivery of the alert item SO_8. In some embodiments, the eighth speech output SO_8 is modified to reflect the delay. For example, the modified speech output is "It's 4 to 2, Time to call mom."

[0201] Continuing with the example, while the user is driving toward the final waypoint X_5, the user provides a fourth speech input SI_4 (*e.g.*, "Call mom now"). The digital assistant evaluates the time to reach the final destination X_5 determines that the user is likely to reach the final destination before the phone call is completed. The digital assistant compares the relative urgency for establishing and maintaining the call and delivery of the final navigation instruction SO_9 ("You have arrived."). The digital assistant concludes that the final navigation instruction is less urgent than the call because the consequence for not providing the final navigation instruction is not severe. The user can very well see the library without being explicitly told that the destination has been reached. As a result, the digital assistant proceeds to make the call and allows the call to persist through the default delivery time for the final navigation instruction. In some embodiments, the digital assistant shifts the delivery of the final navigation direction to before the call is established, and with some modifications such as "You are almost there." or "Destination in 50 feet." In some embodiments, if the digital assistant later detects that the user actually drove past the destination, the digital assistant can interrupt the call and reroute to get the user back to the destination.

[0202] The above example illustrates an example process for handling user-initiated and system-initiated interruptions based on the present context. FIG. 5B is a flow chart for an exemplary process 500 for handling user-initiated and system-initiated interruptions while the digital assistant is delivering navigation directions in accordance with some embodiments.

[0203] In the exemplary process 500, a navigation request is received (502) from a user. A first information provision process is initiated (504) in response to the navigation request, where the first information provision process includes preparing at least a first navigation instruction (*e.g.*, a first turn-by-turn direction) and a second navigation instruction (*e.g.*, a second turn-by-turn direction). The first navigation instruction is delivered (506) to the user at a respective default delivery time associated with the first navigation instruction. After or concurrent with the delivery of the first navigation instruction, an event operable to initiate a second information provision process is detected (508). The second information provision process is initiated (510) in response to detecting the event, where the second information process includes preparing a respective output to be delivered to the user regarding the event. A relative urgency between the second navigation instruction and the output regarding the event is determined (512). Then, the second navigation instruction and the output regarding the event are provided (514) in an order based on the determined relative urgency.

[0204] In some embodiments, the digital assistant determines respective default delivery times for the second navigation instruction and the output regarding the event. In some embodiments, the digital assistant determines whether there is a timing conflict between deliveries of the second navigation instruction and the output regarding the event according to their respective default delivery times.

[0205] In some embodiments, the digital assistant determines the relative urgency between the second navigation instruction and the output regarding the event upon detecting the timing conflict. In some embodiments, the digital assistant does not determine the relative urgency between the second navigation instruction and the output regarding the event unless the timing conflict has been detected. In some embodiments, the digital assistant overrides at least one of the respective default delivery times of the second navigation instruction and the output regarding the event based on the determined relative urgency.

[0206] In some embodiments, the digital assistant receives an information request from the user, and the information request does not modify the directions request.

The information request is the event operable to initiate the second information provision process. In some embodiments, the output regarding the event includes at least a speech output containing information retrieved in response to the information request.

[0207] In some embodiments, the digital assistant detects occurrence of a trigger event for a previously established reminder. The output regarding the event includes at least an alert item providing content of the previously established reminder. In some embodiments, the respective default time for delivering the output regarding the event is a reminder time specified in a previously established reminder.

[0208] In some embodiments, the digital assistant detects arrival of a push-notification from a third-party application or process not currently controlled by the digital assistant. The digital intercepts the push-notification before the push-notification is presented to the user. In some embodiments, the output regarding the event is a speech output prepared by the digital assistant regarding the arrival of the push-notification. In some embodiments, the respective default time for delivering the output regarding the event is immediately after the arrival of the push-notification.

[0209] In some embodiments, providing the second navigation instruction and the output regarding the event in an order based on the determined relative urgency further includes: determining that the second navigation instruction has a higher relative urgency than the output regarding the event; and delivering the second navigation instruction before the output regarding the event.

[0210] In some embodiments, providing the second navigation instruction and the output regarding the event in an order based on the determined relative urgency further includes: determining that the second navigation instruction has a lower relative urgency than the output regarding the event; and delivering the second navigation instruction after the output regarding the event.

[0211] In some embodiments, the respective default time for providing the second navigation instruction is based on a predetermined proximity between a current location of the user and a respective waypoint associated with the second navigation instruction.

[0212] In some embodiments, determining the relative urgency between the second navigation instruction and the output regarding the event further includes determining a present context associated with the user; and determining the relative urgency between the second response and the third response based the present context associated with the user.

[0213] In some embodiments, determining the relative urgency between the second navigation instruction and the output regarding the event further includes determining whether the second navigation instruction is associated with a waypoint within a predetermined distance from a current location of the user.

[0214] In some embodiments, determining the relative urgency between the second navigation instruction and the output regarding the event further includes determining whether the user is likely to pass a waypoint associated with the second navigation instruction within a predetermined time window from a current time. In some embodiments, the predetermined time window is based on a respective time buffer required to deliver the third response to the user.

[0215] In some embodiments, determining the relative urgency between the second navigation instruction and the output regarding the event further includes: upon determining that the user is likely to pass by the waypoint associated with the second navigation instruction within the predetermined time window from the current time, assigning a higher level of urgency to the second navigation instruction than the output regarding the event.

[0216] In some embodiments, determining the relative urgency between the second navigation instruction and the output regarding the event further includes: upon determining that the user is unlikely to pass by the waypoint associated with the second navigation instruction within the predetermined time window from the current time, assigning a higher level of urgency to the second navigation instruction than the output regarding the event.

[0217] In some embodiments, determining the present context associated with the user further includes: monitoring a current location, a current direction, and a current speed of the user (*e.g.*, using a GPS device); and determining the present context based on the monitoring.

[0218] In some embodiments, the output regarding the event is an alert item generated for a previously established reminder or arrival of a third-party notification, and determining the relative urgency further includes: based on the present context, determining whether delivery of the second navigation instruction at the respective default delivery time of the second navigation instruction is likely to adversely affect a utility of the reminder or notification to the user. In some embodiments, upon determining that delivery of the second navigation instruction at the respective default delivery time of the second turn-by-turn direction is likely to affect the utility of the reminder or notification, the digital assistant assigns a lower relative urgency to the second navigation instruction than the alert item, and delays delivery of the second navigation instruction until after delivery of the alert item for the reminder or notification to the user.

[0219] In some embodiments, upon determining that delivery of the second navigation instruction at the respective default delivery time of the second navigation instruction is unlikely to affect the utility of the reminder or notification, the digital assistant assigns a higher relative urgency to the second navigation instruction than the alert item for the reminder or notification, and delays delivery of the alert item for the reminder or notification until after delivery of the second navigation instruction to the user. In some embodiments, the digital assistant determines that the utility of the reminder or notification is unlikely to be affected during a period between a reminder time that is specified in the previously established re-

minder and a threshold amount of preparation time needed before an event time specified in the previously established reminder.

[0220] The above process is described in the context of providing navigation instructions during navigation. Another example information provision process that benefits from the context-sensitive, case-by-case relative urgency evaluation described in this specification is a real-time cooking or experiment instruction provided by the digital assistant. For example, the user can request real-time cooking instructions from the digital assistant by a speech input. The digital assistant reads a series of cooking steps at appropriate times, such that the cooking is completed in the correct amount of time. For example, during the instruction process, the digital assistant provides an instruction "Put pasta in the pot and boil for five minutes." After that instruction, the digital assistant waits 30 seconds (*e.g.*, a time buffer associated with the instruction) for the user to complete the action instructed. The digital assistant then provides the next instruction "Cut the veggies into stripes." The digital assistant remains silent while the user is cutting up the veggies. If the user suddenly says "How to I cut the peppers?" shortly before the expiry of the 5 minutes cooking time. The digital assistant determines whether to respond to the user's question or to provide the next instruction (*e.g.*, "Take the pasta out of the pot now."). If the user had asked his question way before the end of the 5 minutes cooking time, the digital assistant would have decide that it is more urgent to provide the answer to the new question about cutting peppers. But in this scenario, the digital assistant will consider providing the next instruction more urgent because the pasta would over cook if not taken out of the pot at the right time. After the digital assistant has provided the instruction "Take the pasta out of the pot now," the digital assistant evaluates the relative urgency between the answer to the user's veggie-related question and the next instruction in the cooking recipe (*e.g.*, "Put the vegetable stripes in the pot."). In this scenario, the digital assistant considers it is more urgent to answer to user's question about how to cut the veggies into stripes, because the delaying the answer to the user's question will adversely affect the utility of the next instruction to the user. In other words, the next instruction (*e.g.*, "Put the vegetable stripes in the pot.") is not useful to the user, until the user has found out how to cut the veggies first. As a result, the digital assistant answers the user's question regarding how to cut peppers with a series of instructions, and delays the delivery of the next instruction (*e.g.*, "Put the vegetable stripes in the pot.") until after the instructions regarding cutting peppers have been provided to the user.

[0221] Another example information provision process that benefits from the context-sensitive, case-by-case relative urgency evaluation described in this specification is a real-time exercise instruction provided by the digital assistant. For example, the user can request real-time instructions from the digital assistant by a speech input. The digital assistant reads a series of exercise routines at appropriate times, such that the each routine is performed for the correct amount of time. For example, during the instruction process, the digital assistant provides an instruction "Routine 1, Raise your arm and stand still." After that instruction, the digital assistant waits 60 seconds for the user to complete the routine as instructed. Before the end of the routine, the user asks "Find me a good ice cream place." The digital assistant determines whether to respond to the user's search request or to read the next exercise instruction (*e.g.*, "Now put down your arms.") at the end of the 60 seconds. In this case, 60 seconds is too short a time to accommodate the answer to the user's search request, and the answer is not time-critical. Therefore, the digital assistant determines that the next instruction is more urgent and delivers the instruction at the end of 60 seconds. Afterwards, the digital assistant considers whether to read the next instruction ("Now, lift your left leg and stand on your right leg for 5 minutes.") or to provide the answer to the search request (*e.g.*, search for ice cream shops). Again, the digital assistant determines that it is more urgent to provide the next exercise instruction because the answer to the user's question will remain valid until after provision of the next exercise instruction, and searching for an ice cream shop does not appear to be an urgent matter based on the priority parameters established by the digital assistant. After the digital assistant provides the next instruction, and the digital assistant proceeds to provide the search results on ice cream shops to the user through a series of speech outputs. Suddenly, while the digital assistant is in the middle of reading the list of search results to the user, the user shouts "Ouch, I am hurt, call 911!" In this example scenario, the digital assistant maintains a continuous listening mode and captures this speech input from the user. The digital assistant determines whether to continue reading the search results or to respond to the user's new speech input. Based on the content of the new speech input, the digital assistant determines that an emergency has occurred, and the digital assistant immediately suspends the search result reading, and responds to the user "OK, right away." While the digital assistant makes contact with the emergency services, the digital assistant optionally provides a series of self-assistance instructions to the user, such as "Stay still." or "Are you bleeding?" and so on.

[0222] In some embodiments, the digital assistant maintains (*e.g.*, suspends but does not abandon) the information provision process for the search result reading for an extended period of time. For example, in some embodiments, after the user has been treated by the emergency services, and the user seems to return a calm and collected state, the digital assistant resumes the suspended information provision process. In some embodiments, the digital assistant optionally provides a speech output to bring the user back to the context of the suspended information provision process. For example, in this example scenario, the digital assistant provides a

speech output "Still interested in the ice cream shops?" At this point, the user can permit the digital assistant to resume the suspended search result reading, or terminates it by saying "No."

[0223] In accordance with some embodiments, FIG. 6 shows a functional block diagram of an electronic device 600 configured in accordance with the principles of the invention as described above. The functional blocks of the device may be implemented by hardware, software, or a combination of hardware and software to carry out the principles of the invention. It is understood by persons of skill in the art that the functional blocks described in FIG. 6 may be combined or separated into sub-blocks to implement the principles of the invention as described above. Therefore, the description herein may support any possible combination or separation or further definition of the functional blocks described herein.

[0224] As shown in FIG. 6, the electronic device 600 includes a speech receiving unit 602, a first information provision unit 604, a first response provision unit 606, a detection unit 608, a second information provision unit 610, an urgency determining unit 612, and a second response provision unit 614. The speech receiving unit 602 is configured to receive a first speech input from a user. The first information provision unit 604 is configured to initiate a first information provision process in response to receipt of the first speech input, the first information provision process comprising preparing at least a first response and a second response to the first speech input. The first response provision unit 606 is configured to provide the first response to the user. The detection unit 608 is configured to, after or concurrent with the provision of the first response to the user, but before provision of the second response to the user, detect an event operable to initiate a second information provision process. The second information provision unit 610 is configured to initiate the second information provision process in response to detecting the event, the second information provision process comprising preparing at least a third response to the event. The urgency determining unit 612 is configured to determine a relative urgency between the second response and the third response. The second response provision unit 614 is configured to provide one of the second response and the third response to the user in an order based on the determined relative urgency. In some embodiments, the first information provision unit 604 and the second information provision unit 610 are the same information provision unit. In some embodiments, the first response provision unit 606 and the second response provision unit 614 are the same response provision unit. In some embodiments, the electronic device 600 is further configured to perform other functions and processes described above (e.g., through other function unit(s) 616).

[0225] In accordance with some embodiments, FIG. 7 shows a functional block diagram of an electronic device 700 configured in accordance with the principles of the invention as described above. The functional blocks of the device may be implemented by hardware, software, or a combination of hardware and software to carry out the principles of the invention. It is understood by persons of skill in the art that the functional blocks described in FIG. 7 may be combined or separated into sub-blocks to implement the principles of the invention as described above. Therefore, the description herein may support any possible combination or separation or further definition of the functional blocks described herein.

[0226] As shown in FIG. 7, the electronic device 700 includes a request receiving unit 702, a first information provision unit 704, a delivering unit 706, a detection unit 708, a second information provision unit 710, an urgency determining unit 712, and an instruction provision unit 714. The request receiving unit 702 is configured to receive a navigation request from a user. The information provision unit 704 is configured to initiate a first information provision process in response to the navigation request, the first information provision process comprising preparing at least a first navigation instruction and a second navigation instruction. The delivering unit 706 is configured to deliver the first navigation instruction to the user at a respective default delivery time associated with the first navigation instruction. The detection unit 708 is configured to, after or concurrent with the delivery of the first navigation instruction, detect an event operable to initiate a second information provision process. The second information provision unit 710 is configured to initiate the second information provision process in response to detecting the event, the second information process comprising preparing a respective output to be delivered to the user regarding the event. The urgency determining unit 712 is configured to determine a relative urgency between the second navigation instruction and the output regarding the event. The instruction provision unit 714 is configured to provide the second navigation instruction and the output regarding the event in an order based on the determined relative urgency. In some embodiments, the electronic device 700 is further configured to perform other functions and processes described above (e.g., through other function unit(s) 716).

[0227] The foregoing description, for purpose of explanation, has been described with reference to specific embodiments. However, the illustrative discussions above are not intended to be exhaustive or to limit the invention to the precise forms disclosed. Many modifications and variations are possible in view of the above teachings. The embodiments were chosen and described in order to best explain the principles of the invention and its practical applications, to thereby enable others skilled in the art to best utilize the invention and various embodiments with various modifications as are suited to the particular use contemplated.

Claims

1. A method of operating a digital assistant (326), com-

prising:

at a device (300) having one or more processors (304) and memory (302):

receiving (402) a first speech input from a user;
initiating (404) a first information provision process in response to receipt of the first speech input, the first information provision process comprising preparing at least a first response and a second response to the first speech input;
providing (406) the first response to the user;
after or concurrent with the provision of the first response to the user, but before provision of the second response to the user, detecting (408) an event operable to initiate a second information provision process;
initiating (410) the second information provision process in response to detecting the event, the second information provision process comprising preparing at least a third response to the event;
determining whether the digital assistant is currently operating in a hands-free mode or an eyes-free mode, wherein one or more input channels that require the user's hand movement are disabled in the hands-free mode and wherein visual outputs are disabled in the eyes-free mode;

determining (412) the relative urgency between the second response and the third response upon determining that the digital assistant is currently operating the hands-free mode or the eyes-free mode; and

providing (414) one of the second response and the third response to the user in an order based on the determined relative urgency.

2. The method of claim 1, wherein the first speech input is a navigation request and the first response and the second response are two navigation instructions associated with two different waypoints along a route prepared in response to the navigation request.

3. The method of any of claims 1-2, wherein the first speech input is a search request, and the first response and the second response are speech outputs reading two different search results retrieved in response to the search request.

4. The method of any of claims 1-3, wherein the first speech input is a list-reading request, and the first response is a speech output summarizing a list of information items or a subset thereof, or a speech output reading content of at least one of the list of information items.

5. The method of any of claims 1-4, further comprising:

detecting that the user is currently performing one of the actions of: navigating a vehicle, walking, jogging, exercising, and operating a device or application not currently controlled by the digital assistant; and
invoking the hands-free mode or the eyes-free mode upon said detection, wherein one or more input channels that require the user's hand movement are disabled in the hands-free mode and wherein visual outputs are disabled in the eyes-free mode.

6. The method of any of claims 1-5, further comprising:

determining whether the second response and the third response are suitable to be provided in parallel on different output channels of the device; and
determining the relative urgency between the second and the third responses upon determining that the second response and the third response are not suitable to be provided in parallel on different output channels of the device.

7. The method of any of claims 1-6, wherein determining the relative urgency between the second response and the third response further comprises:

determining a present context associated with the user; and
determining the relative urgency between the second response and the third response based the present context associated with the user.

8. The method of any of claims 1-7, wherein determining the relative urgency between the second response and the third response further comprises:

determining whether the second response is associated with a location within a predetermined distance from a current location of the user.

9. The method of any of claims 1-8, wherein the predetermined distance is based on a current speed of the user.

10. The method of any of claims 1-8, wherein determining the relative urgency between the second response and the third response further comprises:

determining whether the user is likely to pass by a location associated with the second response

within a predetermined time window from a current time.

11. The method of any of claims 1-10 wherein the second response is provision of an information item in response to the first speech input, and the third response is an alert item generated for a previously established reminder or arrival of a push notification, and wherein the method further comprises:

based on the present context, determining whether delivery of the second response at a respective default delivery time of the second response is likely to adversely affect a utility of the reminder or push-notification to the user.

12. A non-transitory computer readable medium having instructions stored thereon, the instructions, when executed by one or more processors, cause the processors to perform any of the methods of claims 1-11.

13. A system (300), comprising:

one or more processors (304); and
memory (302) having instructions stored thereon, the instructions, when executed by the one or more processors, cause the processors to perform any of the methods of claims 1-11.

Patentansprüche

1. Verfahren zum Betreiben eines digitalen Assistenten (326), umfassend:

bei einer Vorrichtung (300), die einen oder mehrere Prozessoren (304) und Speicher (302) aufweist:

Empfangen (402) einer ersten Spracheingabe von einem Benutzer;
Initiieren (404) eines ersten Informationsbereitstellungsverfahrens in Antwort auf den Empfang der ersten Spracheingabe, wobei das erste Informationsbereitstellungsverfahren ein Vorbereiten zumindest einer ersten Antwort und einer zweiten Antwort auf die erste Spracheingabe umfasst;
Bereitstellen (406) der ersten Antwort an den Benutzer;
nach oder gleichzeitig mit dem Bereitstellen der ersten Antwort an den Benutzer, aber bevor dem Bereitstellen der zweiten Antwort an den Benutzer, Erkennen (408) eines Ereignisses, welches betriebsbereit ist, um ein zweites Informationsbereitstellungsverfahren zu initiieren;
Initiieren (410) des zweiten Informationsbereitstellungsverfahrens in Antwort auf das Erkennen des Ereignisses, wobei das zweite Informationsbereitstellungsverfahren ein Vorbereiten zumindest einer dritten Antwort auf das Ereignis umfasst;
Bestimmen, ob der digitale Assistent aktuell in einem Freihandmodus oder einem Augen-frei-Modus betrieben wird, wobei ein oder mehrere Eingabekanäle, welche die Handbewegung des Benutzers erfordern, in dem Freihandmodus deaktiviert sind und wobei visuelle Ausgaben in dem Augen-frei-Modus deaktiviert sind;
Bestimmen (412) der relativen Dringlichkeit zwischen der zweiten Antwort und der dritten Antwort beim Bestimmen, dass der digitale Assistent aktuell in dem Freihandmodus oder Augen-frei-Modus betrieben wird; und
Bereitstellen (414) der zweiten Antwort oder der dritten Antwort an den Benutzer, in einer Reihenfolge, die auf der bestimmten relativen Dringlichkeit basiert.

2. Verfahren nach Anspruch 1, wobei die erste Spracheingabe eine Navigationsanfrage ist und die erste Antwort und die zweite Antwort zwei Navigationsanweisungen sind, die mit zwei unterschiedlichen Wegpunkten entlang einer Route verknüpft sind, welche in Antwort auf die Navigationsanfrage vorbereitet wird.

3. Verfahren nach einem der Ansprüche 1 bis 2, wobei die erste Spracheingabe eine Suchanfrage ist und die erste Antwort und die zweite Antwort Sprachausgaben sind, die zwei unterschiedliche Suchresultate lesen, die in Antwort auf die Suchanfrage abgerufen werden.

4. Verfahren nach einem der Ansprüche 1 bis 3, wobei die erste Spracheingabe eine Listenleseanfrage ist und die erste Antwort eine Spracheingabe ist, welche eine Liste von Informationselementen oder eine Untermenge davon zusammenfasst oder eine Sprachausgabe ist, die Inhalt von zumindest eines der Informationselementen der Liste liest.

5. Verfahren nach einem der Ansprüche 1 bis 4, weiterhin umfassend:

Erkennen, dass der Benutzer aktuell eine der folgenden Aktionen durchführt:

Navigieren eines Fahrzeuges, Gehen, Joggen, Turnen und Betreiben einer Vorrichtung oder einer Anwendung, die aktuell nicht durch den digitalen Assistenten kon-

trolliert wird; und
Aufrufen des Freihandmodus oder des Augen-frei-Modus bei der Erkennung, wobei ein oder mehrere Eingabekanäle, welche eine Handbewegung des Benutzers benötigen in dem Freihandmodus deaktiviert sind und wobei visuelle Ausgaben in dem Augen-frei-Modus deaktiviert sind.

6. Verfahren nach einem der Ansprüche 1 bis 5, weiterhin umfassend:

Bestimmen, ob die zweite Antwort und die dritte Antwort geeignet sind, um parallel auf unterschiedlichen Ausgabekanälen der Vorrichtung bereitgestellt zu werden; und
Bestimmen der relativen Dringlichkeit zwischen der zweiten und der dritten Antwort beim Bestimmen, dass die zweite Antwort und die dritte Antwort nicht geeignet sind, um parallel auf unterschiedlichen Ausgabekanälen der Vorrichtung bereitgestellt zu werden.

7. Verfahren nach einem der Ansprüche 1 bis 6, wobei das Bestimmen der relativen Dringlichkeit zwischen der zweiten Antwort und der dritten Antwort weiterhin umfasst:

Bestimmen eines aktuellen Zusammenhangs, der mit dem Benutzer verknüpft ist; und
Bestimmen der relativen Dringlichkeit zwischen der zweiten Antwort und der dritten Antwort basierend auf dem aktuellen Kontext, mit dem der Benutzer verknüpft ist.

8. Verfahren nach einem der Ansprüche 1 bis 7, wobei das Bestimmen der relativen Dringlichkeit zwischen der zweiten Antwort und der dritten Antwort weiterhin umfasst:

Bestimmen, ob die zweite Antwort mit einem Standort innerhalb einer vorbestimmten Distanz von einem aktuellen Standort des Benutzers verknüpft ist.

9. Verfahren nach einem der Ansprüche 1 bis 8, wobei die vorbestimmte Distanz auf einer aktuellen Geschwindigkeit des Benutzers basiert.

10. Verfahren nach einem der Ansprüche 1 bis 8, wobei das Bestimmen der relativen Dringlichkeit zwischen der zweiten Antwort und der dritten Antwort weiterhin umfasst:

Bestimmen, ob der Benutzer wahrscheinlich an einem Standort, der mit der zweiten Antwort innerhalb eines vordefinierten Zeitfensters von einer aktuellen Zeit verknüpft ist, vorbeigeht.

11. Verfahren nach einem der Ansprüche 1 bis 10, wobei die zweite Antwort eine Bereitstellung eines Informationselementes in Antwort auf die erste Spracheingabe ist und die dritte Antwort ein Alarmelement ist, welches für eine vorgängig eingerichtete Erinnerung oder eine Ankunft einer Push-Nachricht erzeugt wird und wobei das Verfahren weiterhin umfasst:

basierend auf dem aktuellen Zusammenhang, Bestimmen, ob eine Lieferung der zweiten Antwort zu einer jeweiligen vorgegebenen Lieferzeit der zweiten Antwort wahrscheinlich eine Verwendbarkeit der Erinnerungs- oder Push-Nachricht an den Benutzer negativ beeinflusst.

12. Nichtflüchtiges computerlesbares Medium, welches darauf gespeicherte Anweisungen aufweist, wobei die Anweisungen, wenn sie durch einen oder durch mehrere Prozessoren ausgeführt werden, den Prozessor dazu veranlassen irgendeines der Verfahren nach Anspruch 1 bis 11 durchzuführen.

13. System (300), umfassend:

einen oder mehrere Prozessoren (304); und
Speicher (302), der darauf gespeicherte Anweisungen aufweist, wobei die Anweisungen, wenn sie durch den einen oder die mehreren Prozessoren ausgeführt werden, den Prozessor dazu veranlassen irgendeines der Verfahren nach Anspruch 1 bis 11 durchzuführen.

Revendications

1. Un procédé de mise en oeuvre d'un assistant numérique (326), comprenant :

au niveau d'un dispositif (300) possédant un ou plusieurs processeurs (304) et une mémoire (302) :

la réception (402) d'une première entrée vocale en provenance d'un utilisateur ;
le déclenchement (404) d'un premier processus de délivrance d'information en réponse à la réception de la première entrée vocale, le premier processus de délivrance d'information comprenant la préparation d'au moins une première réponse et d'une seconde réponse à la première entrée vocale ;
la délivrance (406) de la première réponse à l'utilisateur ;
après, ou concurremment à, la délivrance de la première réponse à l'utilisateur, mais avant la délivrance de la seconde réponse

à l'utilisateur, la détection (408) d'un évènement pouvant être mis en oeuvre pour déclencher un second processus de délivrance d'information ;
le déclenchement (410) du second processus de délivrance d'information en réponse à la détection de l'évènement, le second processus de détection d'information comprenant la préparation d'au moins une troisième réponse à l'évènement ;
la détermination si l'assistant numérique est en train de fonctionner dans un mode mains libres ou un mode à l'aveugle ou non, un ou plusieurs canaux nécessitant le mouvement de la main de l'utilisateur étant désactivés dans le mode mains libres et des sorties visuelles étant désactivées dans le mode à l'aveugle ;
la détermination (412) de l'urgence relative entre la seconde réponse et la troisième réponse sur détermination que l'assistant numérique est en train de fonctionner dans le mode mains libres ou le mode à l'aveugle ; et
la délivrance (414) à l'utilisateur de l'une d'entre la seconde réponse et la troisième réponse dans un ordre basé sur l'urgence relative déterminée.

2. Le procédé de la revendication 1, dans lequel la première entrée vocale est une requête de navigation et la première réponse et la seconde réponse sont deux instructions de navigation associées à deux points de repère différents le long d'un itinéraire préparé en réponse à la requête de navigation.

3. Le procédé de l'une des revendications 1 à 2, dans lequel la première entrée vocale est une requête de recherche, et la première réponse et la seconde réponse sont des sorties vocales lisant deux résultats de recherche différents récupérés en réponse à la requête de recherche.

4. Le procédé de l'une des revendications 1 à 3, dans lequel la première entrée vocale est une requête de lecture de liste et la première réponse est une sortie vocale récapitulant une liste d'éléments d'information ou un sous-ensemble de celle-ci, ou une sortie vocale lisant le contenu d'au moins l'un des éléments de la liste d'éléments d'information.

5. Le procédé de l'une des revendications 1 à 4, comprenant en outre :

la détection que l'utilisateur est en train d'effectuer l'une des actions suivantes : conduite d'un véhicule, marche, course à pied, exercice et actionnement d'un dispositif ou d'une application non actuellement contrôlé par l'assistant numérique ; et
appel du mode mains libres ou du mode à l'aveugle suite à ladite détection, un ou plus des canaux d'entrée qui nécessitent le mouvement de la main de l'utilisateur étant désactivés dans le mode mains libres et les sorties visuelles étant désactivées dans le mode à l'aveugle.

6. Le procédé de l'une des revendications 1 à 5, comprenant en outre :

la détermination si la seconde réponse et la troisième réponse sont ou non susceptibles d'être délivrées en parallèle sur des canaux de sortie différents du dispositif ; et
la détermination d'une urgence relative entre la seconde et la troisième réponse sur détermination que la seconde réponse et la troisième réponse ne sont pas susceptibles d'être délivrées en parallèle sur des canaux de sortie différents du dispositif.

7. Le procédé de l'une des revendications 1 à 6, dans lequel la détermination de l'urgence relative entre la seconde réponse et la troisième réponse comprend en outre :

la détermination d'un contexte actuel associé à l'utilisateur ; et
la détermination de l'urgence relative entre la seconde réponse et la troisième réponse sur la base du contexte actuel associé à l'utilisateur.

8. Le procédé de l'une des revendications 1 à 7, dans lequel la détermination de l'urgence relative entre la seconde réponse et la troisième réponse comprend en outre :

la détermination si la seconde réponse est ou non associée à un emplacement situé dans les limites d'une distance prédéterminée par rapport à un emplacement actuel de l'utilisateur.

9. Le procédé de l'une des revendications 1 à 8, dans lequel la distance prédéterminée est basée sur une vitesse actuelle de l'utilisateur.

10. Le procédé de l'une des revendications 1 à 8, dans lequel la détermination de l'urgence relative entre la seconde réponse et la troisième réponse comprend en outre :

la détermination si l'utilisateur est ou non susceptible de passer par un emplacement associé à la seconde réponse dans les limites d'une fenêtre temporelle prédéterminée par rapport à un instant actuel.

11. Le procédé de l'une des revendications 1 à 10, dans lequel la seconde réponse est la délivrance d'un élément d'information en réponse à la première entrée vocale, et la troisième réponse est un élément d'alerte généré pour un rappel précédemment établi ou l'arrivée d'une notification *push*, et dans lequel le procédé comprend en outre :

sur la base du contexte actuel, la détermination si la délivrance de la seconde réponse à un instant de délivrance par défaut respectif de la seconde réponse est susceptible d'affecter de façon négative une utilité du rappel ou de la notification *push* pour l'utilisateur.

12. Un support non transitoire lisible par calculateur sur lequel sont stockées des instructions, les instructions, lorsqu'elles sont exécutées par un ou plusieurs processeurs, faisant en sorte que les processeurs mettent en oeuvre l'un quelconque des procédés des revendications 1 à 11.

13. Un système (300), comprenant :

un ou plusieurs processeurs (304) ; et
une mémoire (302) sur laquelle sont stockées des instructions, les instructions, lorsqu'elles sont exécutées par les un ou plusieurs processeurs, faisant en sorte que les processeurs mettent en oeuvre l'un quelconque des procédés des revendications 1 à 11.

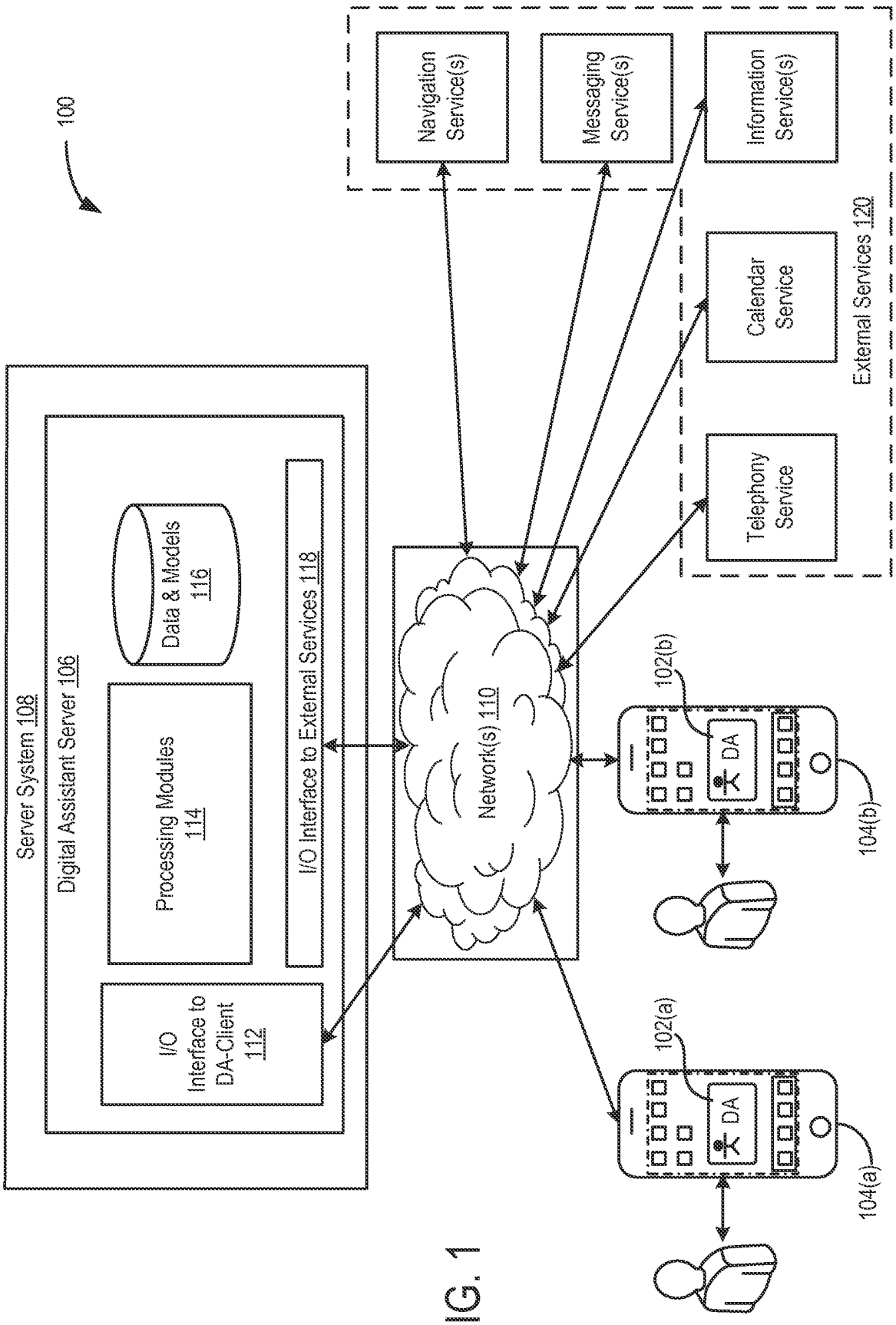

100
Server System 108
Digital Assistant Server 106
Processing Modules 114
Data & Models 116
I/O Interface to External Services 118
I/O Interface to DA-Client 112
Network(s) 110
External Services 120
Navigation Service(s)
Messaging Service(s)
Information Service(s)
Calendar Service
Telephony Service
102(a)
DA
104(a)
102(b)
DA
104(b)

FIG. 1

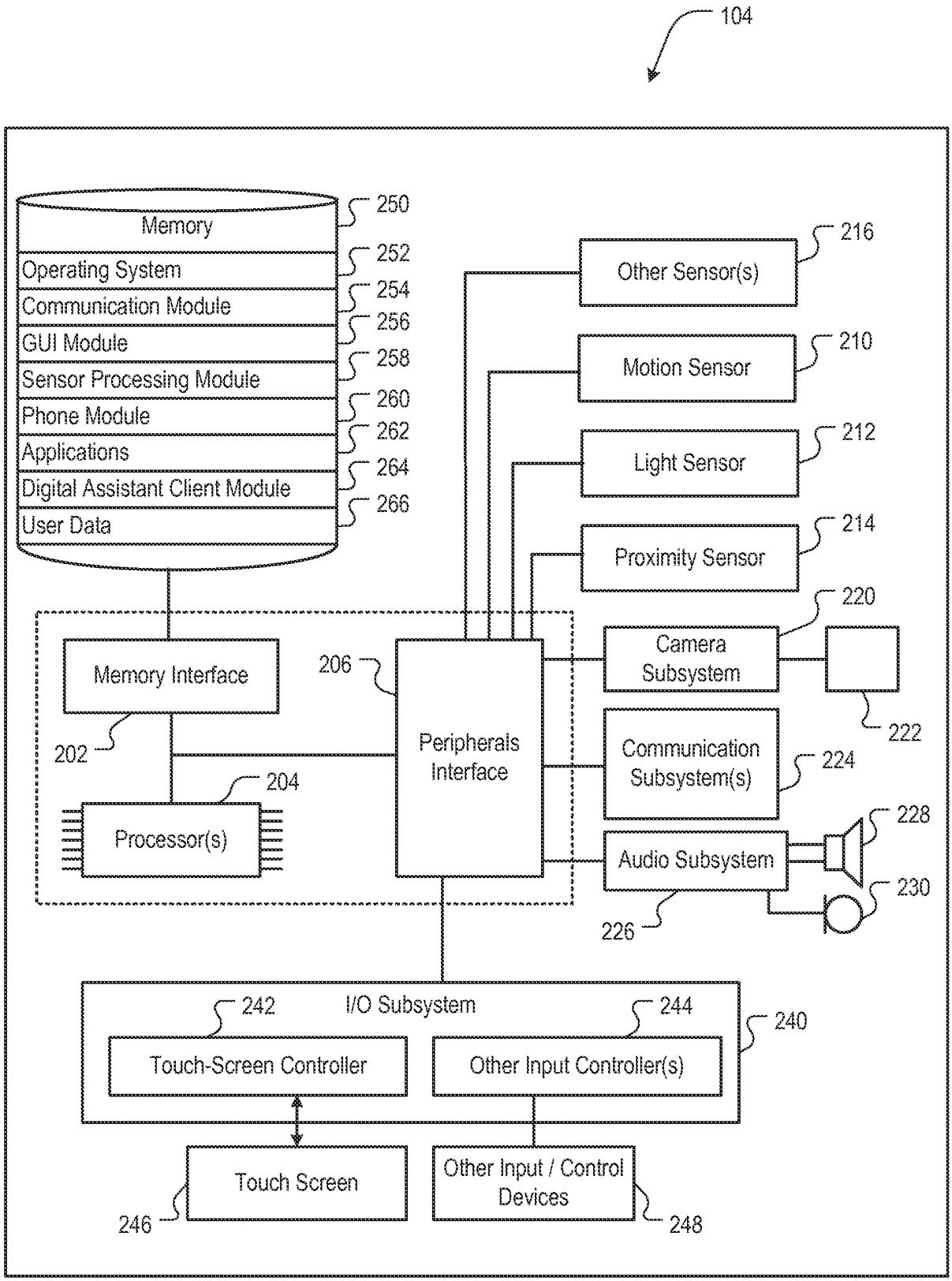

104
Memory
250
Operating System
252
Communication Module
254
GUI Module
256
Sensor Processing Module
258
Phone Module
260
Applications
262
Digital Assistant Client Module
264
User Data
266
Other Sensor(s)
216
Motion Sensor
210
Light Sensor
212
Proximity Sensor
214
Memory Interface
202
206
Peripherals Interface
Camera Subsystem
220
222
Communication Subsystem(s)
224
Audio Subsystem
226
228
230
Processor(s)
204
I/O Subsystem
240
242
244
Touch-Screen Controller
Other Input Controller(s)
Touch Screen
246
Other Input / Control Devices
248

FIG. 2

FIG. 3A

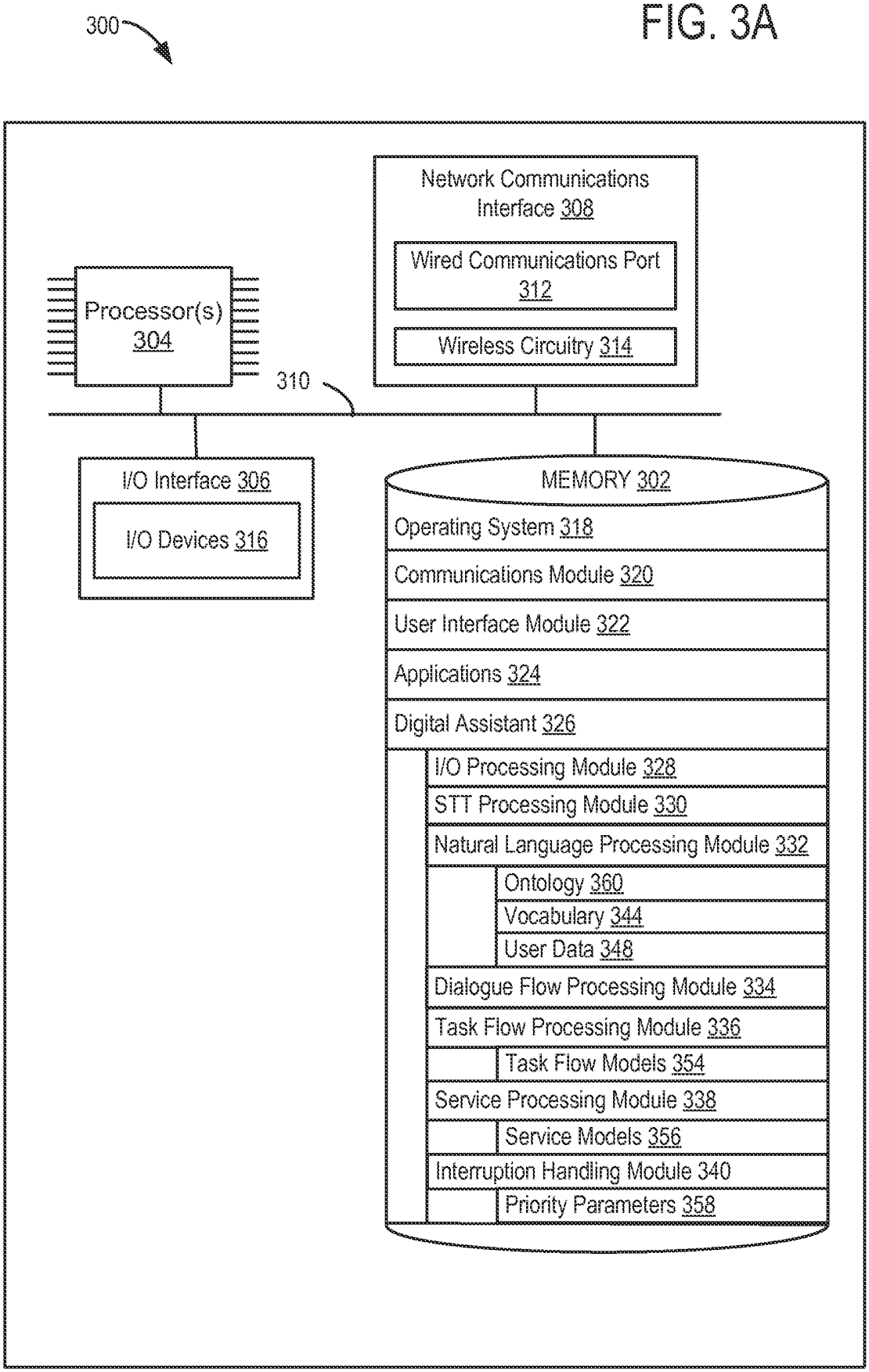
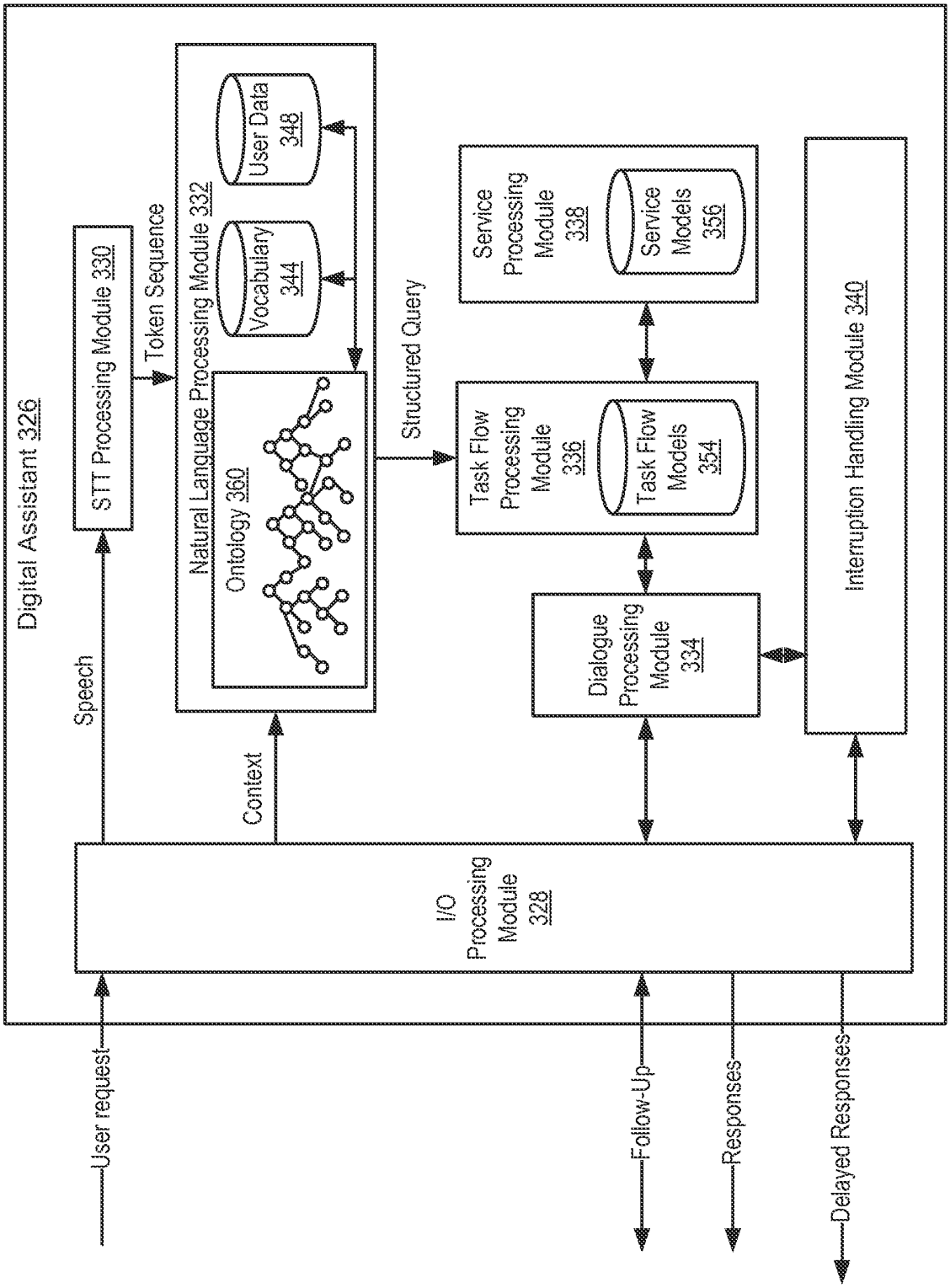

300
Network Communications Interface 308
Wired Communications Port 312
Wireless Circuitry 314
Processor(s) 304
310
I/O Interface 306
I/O Devices 316
MEMORY 302
Operating System 318
Communications Module 320
User Interface Module 322
Applications 324
Digital Assistant 326
I/O Processing Module 328
STT Processing Module 330
Natural Language Processing Module 332
Ontology 360
Vocabulary 344
User Data 348
Dialogue Flow Processing Module 334
Task Flow Processing Module 336
Task Flow Models 354
Service Processing Module 338
Service Models 356
Interruption Handling Module 340
Priority Parameters 358

FIG. 3B


Digital Assistant 326
STT Processing Module 330
Token Sequence
Natural Language Processing Module 332
Ontology 360
Vocabulary 344
User Data 348
Structured Query
Service Processing Module 338
Service Models 356
Task Flow Processing Module 336
Task Flow Models 354
Interruption Handling Module 340
Dialogue Processing Module 334
Speech
Context
I/O Processing Module 328
User request
Follow-Up
Responses
Delayed Responses

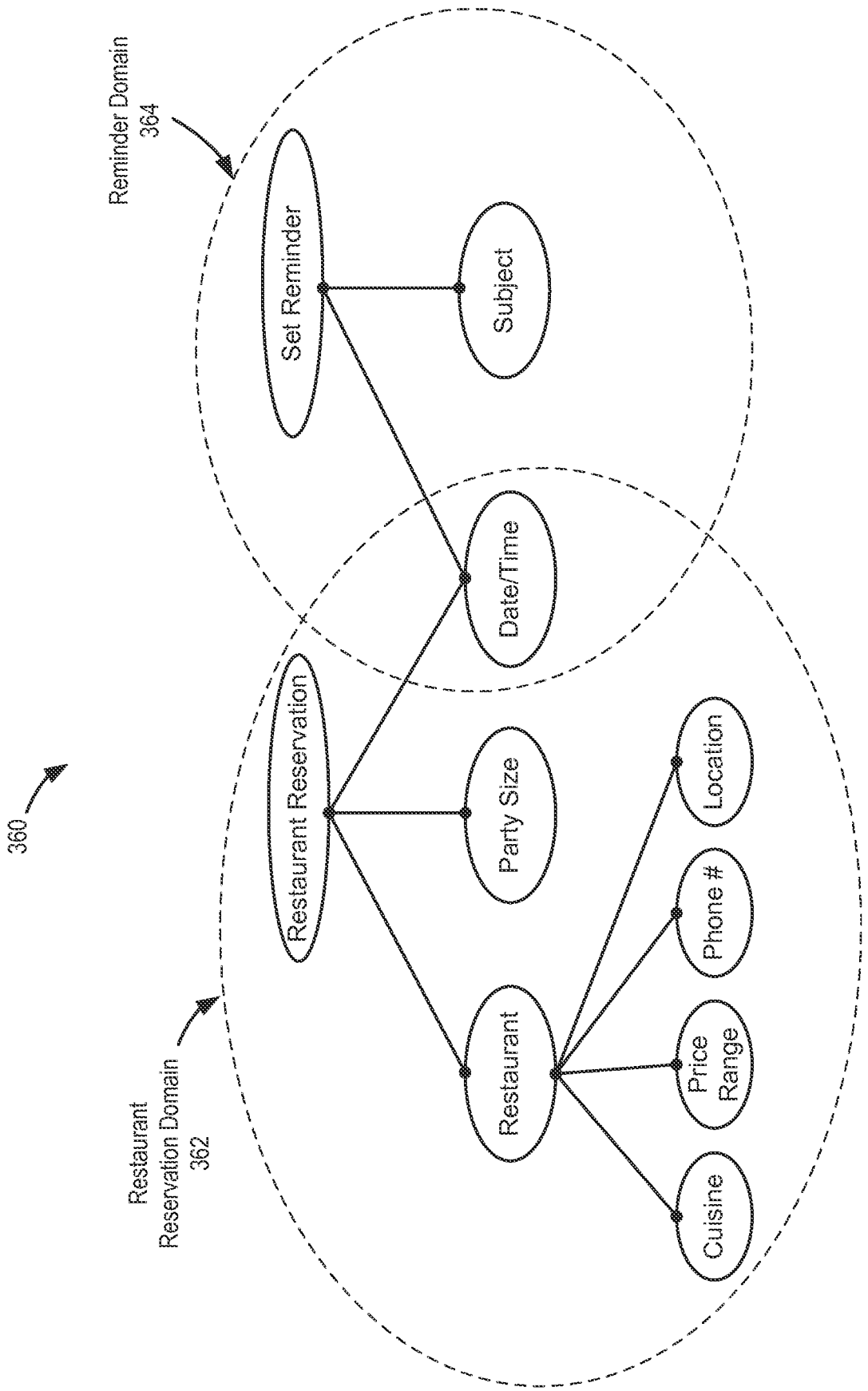

360
Reminder Domain
364
Set Reminder
Subject
Date/Time
Restaurant Reservation
Party Size
Restaurant
Location
Phone #
Price Range
Cuisine
Restaurant Reservation Domain
362

FIG. 3C

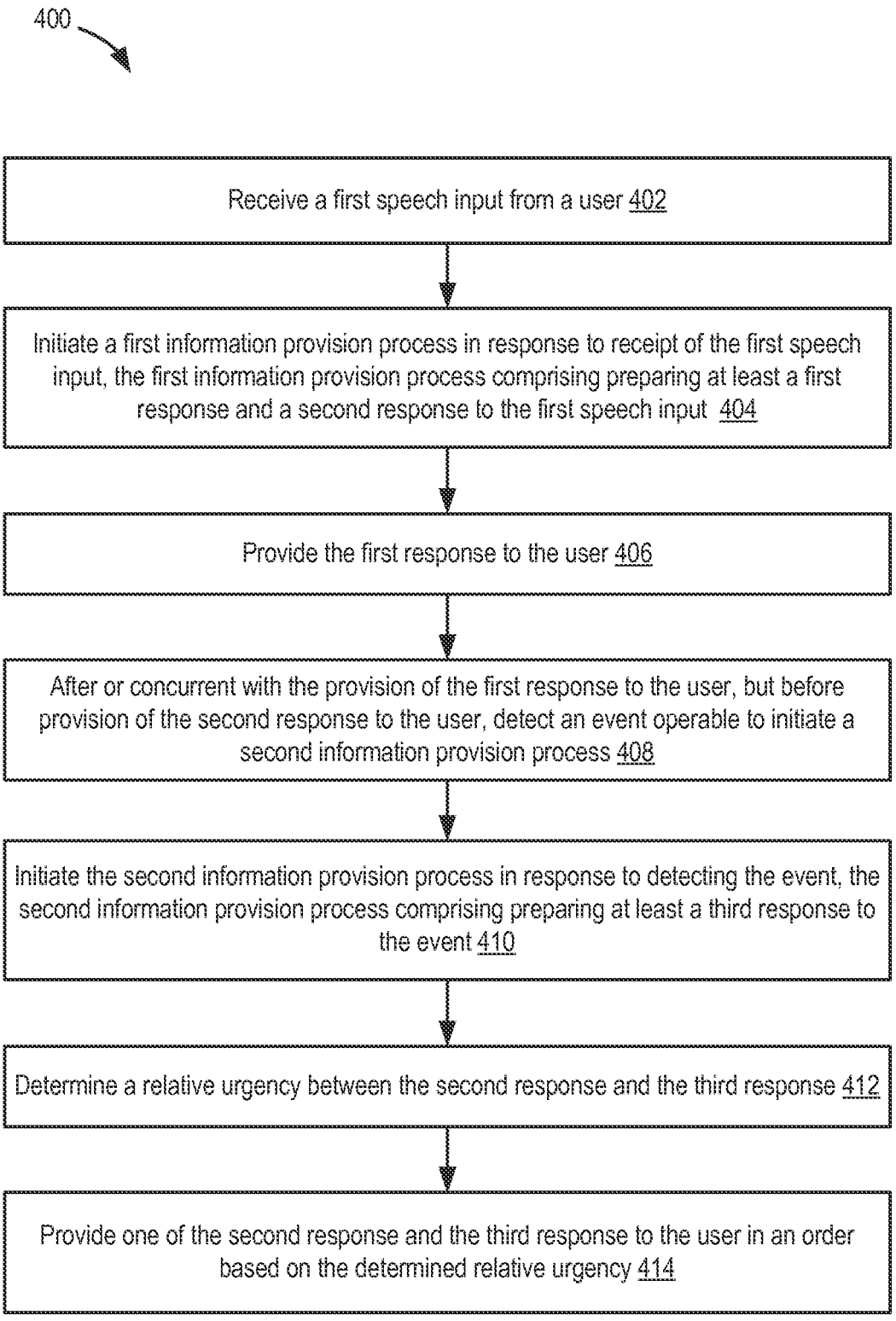

400
Receive a first speech input from a user 402
Initiate a first information provision process in response to receipt of the first speech input, the first information provision process comprising preparing at least a first response and a second response to the first speech input 404
Provide the first response to the user 406
After or concurrent with the provision of the first response to the user, but before provision of the second response to the user, detect an event operable to initiate a second information provision process 408
Initiate the second information provision process in response to detecting the event, the second information provision process comprising preparing at least a third response to the event 410
Determine a relative urgency between the second response and the third response 412
Provide one of the second response and the third response to the user in an order based on the determined relative urgency 414

FIG. 4

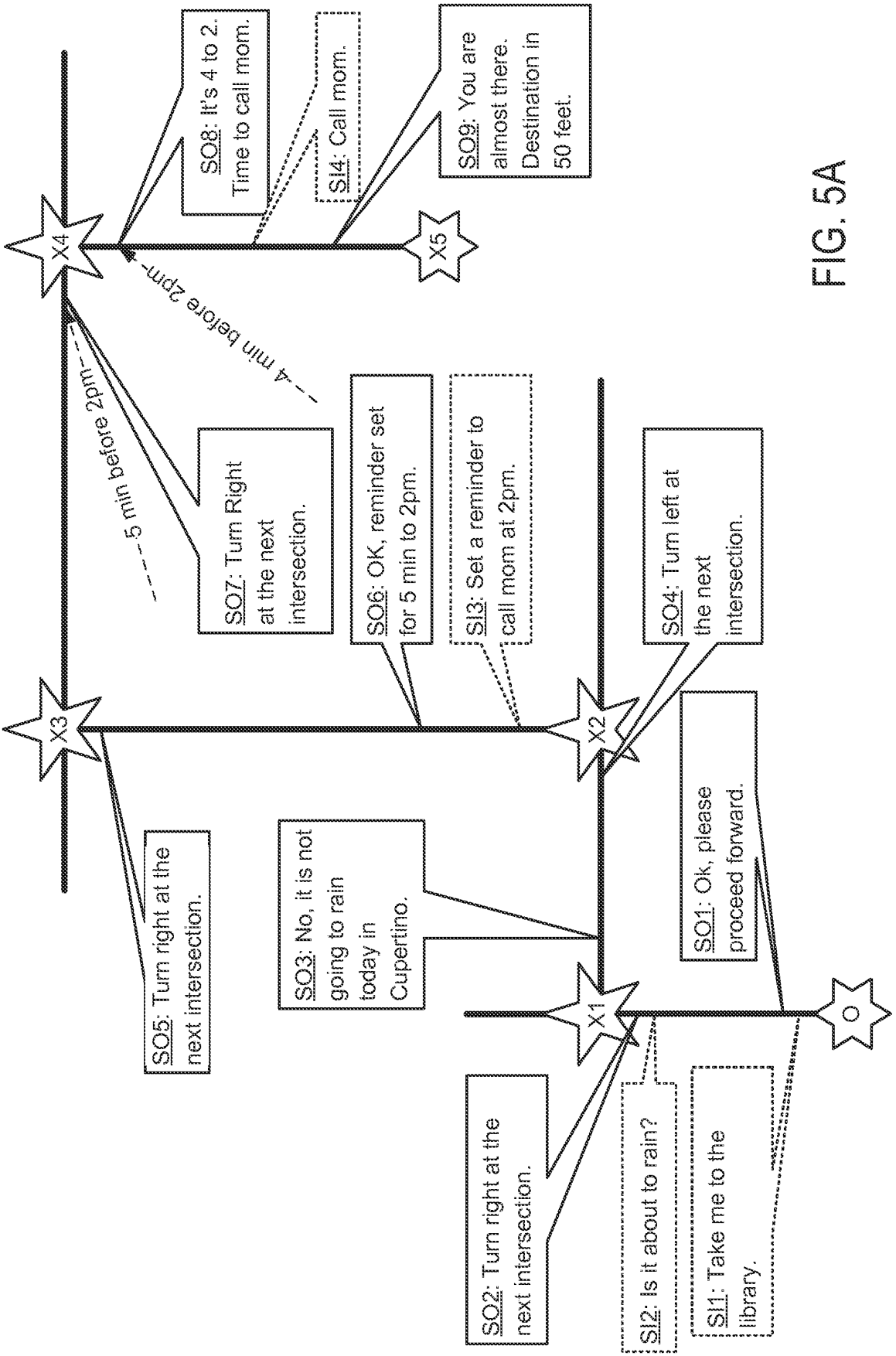

X1
X2
X3
X4
X5
O
SO1: Ok, please proceed forward.
SO2: Turn right at the next intersection.
SO3: No, it is not going to rain today in Cupertino.
SO4: Turn left at the next intersection.
SO5: Turn right at the next intersection.
SO6: OK, reminder set for 5 min to 2pm.
SO7: Turn Right at the next intersection.
SO8: It's 4 to 2. Time to call mom.
SO9: You are almost there. Destination in 50 feet.
SI1: Take me to the library.
SI2: Is it about to rain?
SI3: Set a reminder to call mom at 2pm.
SI4: Call mom.
5 min before 2pm
4 min before 2pm

FIG. 5A

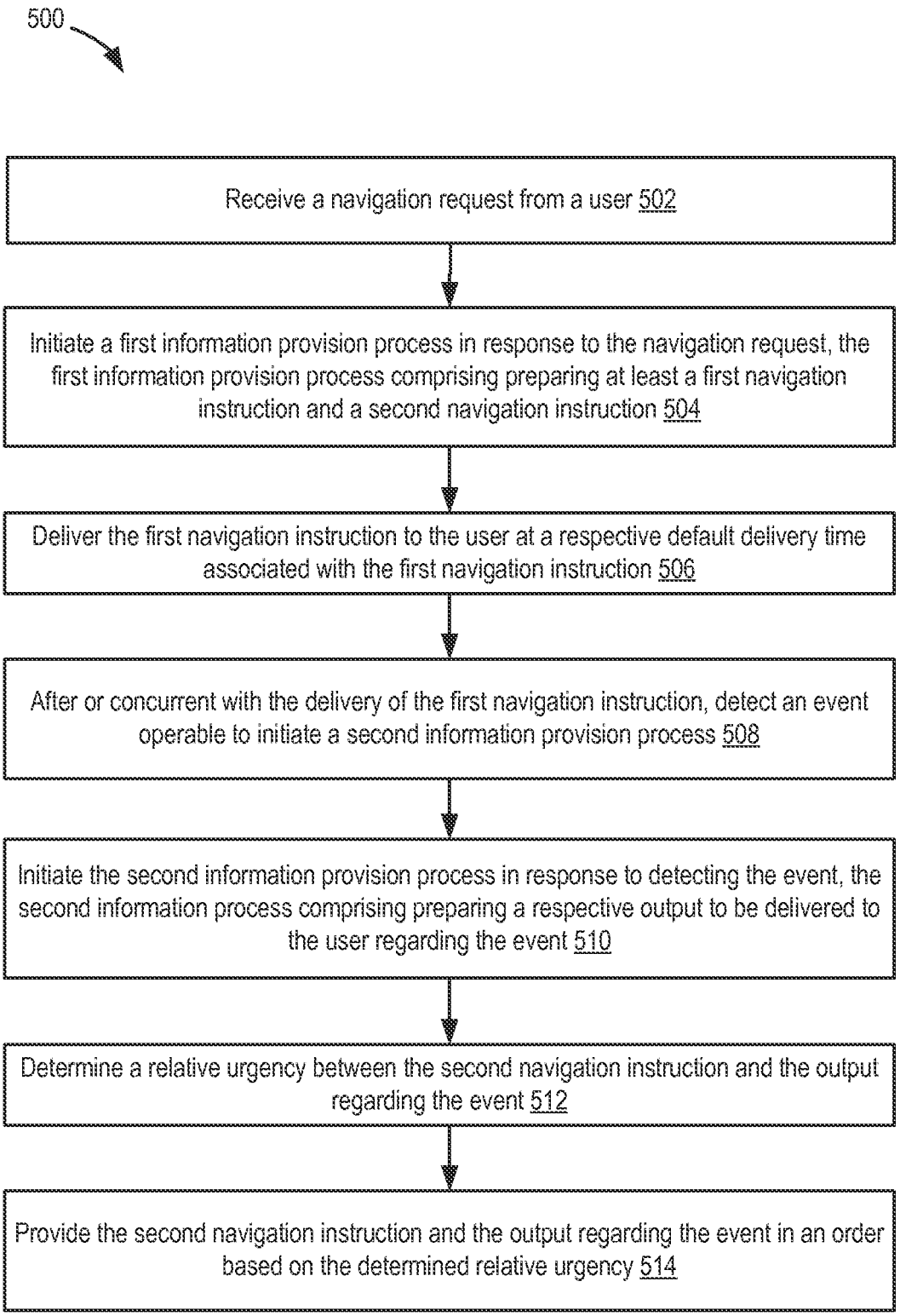

500
Receive a navigation request from a user 502
Initiate a first information provision process in response to the navigation request, the first information provision process comprising preparing at least a first navigation instruction and a second navigation instruction 504
Deliver the first navigation instruction to the user at a respective default delivery time associated with the first navigation instruction 506
After or concurrent with the delivery of the first navigation instruction, detect an event operable to initiate a second information provision process 508
Initiate the second information provision process in response to detecting the event, the second information process comprising preparing a respective output to be delivered to the user regarding the event 510
Determine a relative urgency between the second navigation instruction and the output regarding the event 512
Provide the second navigation instruction and the output regarding the event in an order based on the determined relative urgency 514

FIG. 5B

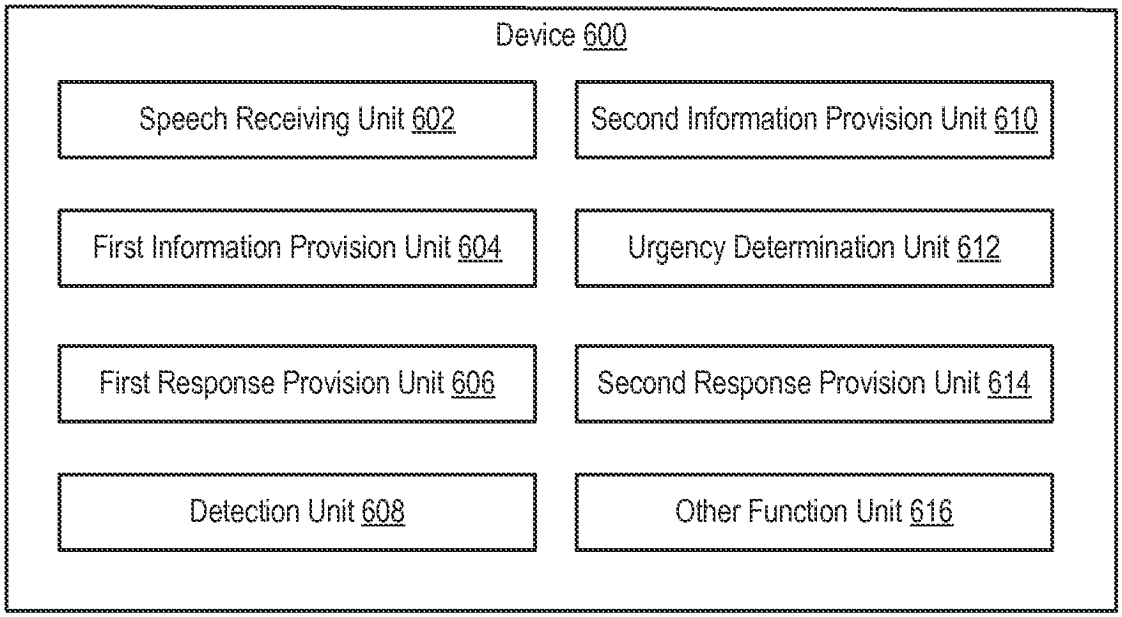

Device 600
Speech Receiving Unit 602
Second Information Provision Unit 610
First Information Provision Unit 604
Urgency Determination Unit 612
First Response Provision Unit 606
Second Response Provision Unit 614
Detection Unit 608
Other Function Unit 616

FIG. 6

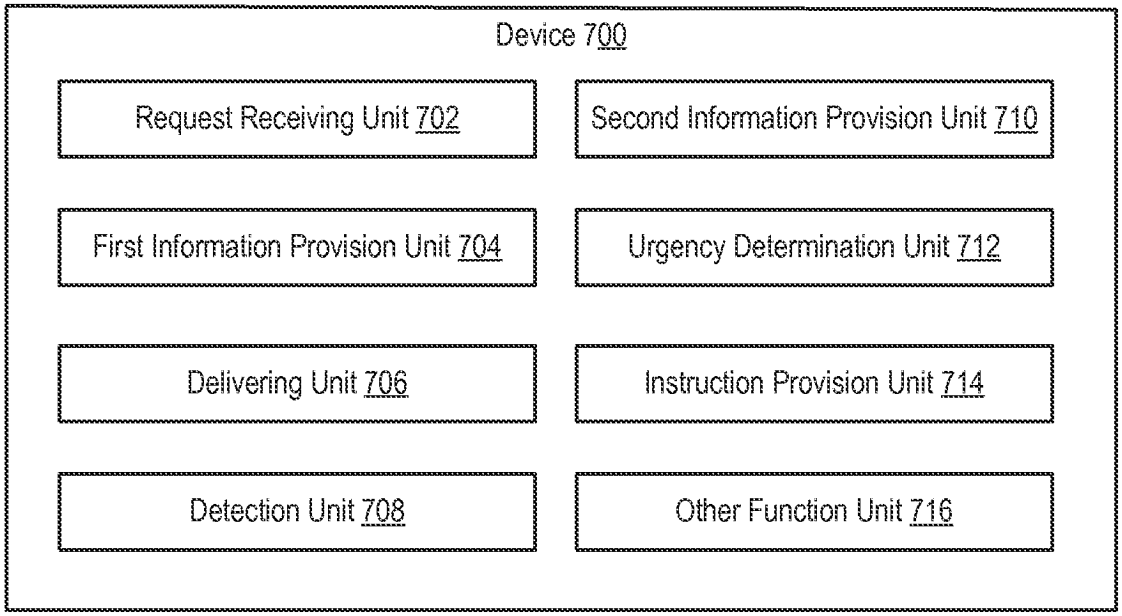

Device 700
Request Receiving Unit 702
Second Information Provision Unit 710
First Information Provision Unit 704
Urgency Determination Unit 712
Delivering Unit 706
Instruction Provision Unit 714
Detection Unit 708
Other Function Unit 716

FIG. 7

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- US 20080077310 A1 **[0004]**
- US 98798211 A **[0075]**
- US 23694211 A **[0108]**
- US 34174308 A **[0122]**
- US 98798210 A **[0132]**
- US 61493201 A **[0132]**